

Benefits of federation

A new association of European neuroscientists highlights the opportunities to establish a transatlantic balance of disciplinary activity, and the failure of others to achieve it.

At their worst, learned societies are dismal, tiny, unrepresentative bodies. In such circumstances, they fail to win the confidence of researchers in their respective discipline in the two activities by which they can in principle best distinguish themselves: their meetings and their publications. At their best, on the other hand, learned societies win allegiance and catalyse coherence through conferences and publications that are lively, authoritative and widely attended or read. That success, especially if enhanced by strong profits from society journals, leaves them well placed to support the professional development of members and also, through their ability to advocate on behalf of a large national proportion of a particular scientific discipline, achieve clout in science politics.

Many US societies have that stature. Regarded enviously in other countries, those such as the American Society for Neuroscience, American Geophysical Union and American Physical Society provide services that attract strong membership from other countries — a feat that even the best of their equivalent bodies abroad find difficult, to say the least.

But who are the equivalents elsewhere? In Europe the answer is, as usual, less than straightforward. The lack of international clout of societies in individual countries and the pursuit of a European vision have led researchers to form societies at the European level, with decidedly mixed and — in comparison with the United States — typically lacklustre results. The success of neuroscientists in setting up a strong association (see page 614) is therefore good news. It also represents a model that other European ventures would do well to note.

One strength of the new body is that national societies have proved willing to modify their activities significantly in order to give the European Neuroscience Association space in which to fulfil itself. Just as importantly, the society is a federation of the national societies. Some European societies have adopted a mixed approach, simultaneously encouraging individual membership while also serving (and taking money from) the national bodies. That approach leads to the tempta-

tion to provide services, such as members' magazines, that are intended to assert the identity and role of the society. Unfortunately, they end up costing a high proportion of the society's budget while achieving poor results due to inadequate resources. Thus the society and its members get the worst possible deal. Much better, if it can, for a European society to publish a revenue-earning journal, if necessary replacing several national journals in the process.

By such demarcation, strong, representative national societies can then be left to do what they do best — foster young scientists' development at national meetings, and represent their disciplines at the national level. Against the wishes of many smaller countries, Europe is stuck fast in a stable if undesirable situation lacking any coherent policy of transnational coordination in basic and strategic research — despite (or partly because of) the European Commission. National funding agencies jealously preserve their independence. In the circumstances, national societies' roles in promoting their members' interests continue to be important.

But whether the national societies like it or not, US societies prefer to operate with a partner at the European level, as do international agencies. Through its meetings and through its institutional contacts, a strong European society can foster international collaboration within and beyond Europe, and can generate support for researchers in poorer European countries.

In short, as long as there are some strong national societies in its discipline in Europe, a federated European society allows the community to get the best at both national and international levels. Those scientists, and those European societies, who strive rather towards an institution representing researchers as individual Europeans are living in a dream world in which there are European research councils, easy mobility of benefits and pensions, and harmonization of education and professional qualifications. Europe's science would do better in such circumstances, and strong European societies can only help hasten that day, but the vision is still a distant one. □

Keep dual funding

The temptation to eliminate Britain's 'dual support system' for university research should be resisted.

In about four weeks' time, British scientists will discover just how deep is the commitment of the Labour government to a flourishing science base. That is when it is due to announce the outcome of its Comprehensive Spending Review, an exercise which, it is now clear, will set the boundaries of public spending over the next three years. There are grounds for optimism. A pre-election commitment to place science "at the heart of government" remains on ministers' lips. Increased spending on research would lie well with the often-repeated promise by Prime Minister Tony Blair that his top spending priority, after health and education, is investment for the future. Chancellor Gordon Brown has promised a slight relaxation in overall spending constraints. And officials from the Office of Science and Technology are reported to have presented a comprehensively documented case for spending an extra £500 million a year on the science base.

But there is a cloud on the horizon which many thought had dissipated, but which reappeared last week. It took the form of reports that Treasury officials are once again canvassing reactions to the proposal for a substantial transfer of public research funds from the four regional university funding councils to the research councils. The apparent logic behind this proposal is clear: research councils are better placed to target research funding to those groups in universities able to use them most productively.

But to follow it would be a recipe for short-term management of university departments in efforts to secure funds at the expense of longer-term research planning. The 'dual support system' may not be ideal, and one should require, as the Dearing committee proposed last year, that research councils pay a fairer proportion of the costs of the research they sponsor. But to abolish the system will undermine, not strengthen, the UK's research base. □



Clinton's ocean agenda offers modest treasures for science

[WASHINGTON] President Bill Clinton last week proposed a broad \$224 million package to protect and restore ocean resources — although it included only modest increases in funding for oceanographic research.

Speaking at a two-day National Ocean Conference in Monterey, California, Clinton and Vice-President Al Gore promised to extend a ban on offshore oil drilling to cover most of the US coastline and to set up an \$800 million trust fund to modernize US ports. They pledged to strengthen protection of coral reefs and to create a permanent oceans commission to coordinate federal policy.

Most of the money — \$194 million — would go into a five-year programme to rebuild and sustain marine fisheries, including the acquisition of three research vessels to provide better data on fish stocks.

An additional \$4 million a year would be spent between 2000 and 2002 on a variety of exploration and research projects. These would include building two new unmanned deep-sea observatories on the Juan de Fuca ridge off the US west coast and in the Gulf of Mexico, and expanding existing shallow-water observatories in Florida and on the continental shelf off New Jersey.



A drop in the ocean: Clinton (front) and Gore pledge little to further these students' research.

The plans also include developing, in partnership with industry, two new submersible vehicles, which will be leased by the government and made available to members of the scientific community.

Another \$12 million would be allocated to the National Oceanic and Atmospheric Administration (NOAA), in order to place hundreds of instrumented buoys in the North Atlantic and North Pacific, beginning in 2000, as part of an expanded Global Ocean Observing System.

Clinton also promised to help speed up the release of classified military data, including underwater acoustical data from listening devices used to track submarines.

Some of the proposals merely reflect plans already under way in NOAA and other federal agencies, while others were still sketchy in their details.

Before the conference, agencies involved in ocean research were canvassed for ideas for future initiatives, which were then screened by the White House Office of Science and Technology Policy and Council on Environmental Quality before being sent on to Clinton and Gore.

But there was little consultation with the ocean science community, says Richard Spinrad, research director for the Consortium for Oceanographic Research and Education (CORE) in Washington.

Spinrad says the research component of the package unveiled in Monterey is "at best a good start" and was nowhere near the kind of investment needed to address the current shortage of observational data for the oceans.

CORE has argued that ocean sciences are chronically underfunded (see *Nature* 379, 283; 1996). Federal spending on oceanographic research has remained flat at about \$600 million a year (in constant dollars) for the past 15 years, while funding for research in general has doubled.

Realizing the research agenda outlined in an influential 1969 report by the Stratton Commission, which led to the creation of NOAA, would require about \$1 billion annually, says Spinrad. Still, he believes that the proposed oceans commission could provide a useful forum for ocean scientists to convey their needs.

Clinton last week asked his cabinet to produce recommendations for a coordinated oceans policy within a year, and said he plans to work with Congress to create the permanent commission.

Even if the initiatives put forth in Monterey were a disappointment, Spinrad takes heart that at least the president and vice-president acknowledged the need for more observational data. "If that message came across during the conference to the Administration and Congress, then that's a great message," he says.

Tony Reichardt

Call to drop charges in French blood affair

[PARIS] The French attorney general has called for all charges to be dropped against three former ministers accused of "collusion in poisoning" (see *Nature* 364, 269; 1993).

Former prime minister Laurent Fabius has been accused, with former deputy health minister Edmond Hervé and former social affairs minister Georgina Dufoix, of delaying the introduction of systematic screening of donated blood for HIV in 1985. He was said to have done this to protect the national market for a French AIDS test, marketed by Diagnostics Pasteur, when a US test from Abbott Laboratories was already available (see *Nature* 367, 673; 1994 & 371, 369; 1994).

But attorney general Jean-François Burgelin concludes that screening, introduced on 19 June 1985, began within a reasonable time, and ahead of many other countries, such as the United Kingdom, which started screening in autumn 1985. He praises Fabius for acting "decisively" in the absence of a clear consensus from the medical and scientific communities. The report criticizes their "passivity" in the face of the emergence of AIDS, and their failure to alert the government.

Hervé is criticized in the brief for

"blindness" and for having failed to take sufficient initiative. But Burgelin concludes that this did not constitute a crime.

The brief also challenges the claim that Abbott was in a position to supply the French market following approval of its test by the US Food and Drug Administration (FDA) in March 1985. French investigators who visited Abbott's Chicago headquarters last year say the company could not provide evidence that this was the case, having destroyed archives relating to the period in line with its policy on expiration of archives.

Other documents raise doubts about the reliability of the Abbott test. A 1986 memo from the American Red Cross shows that the FDA had been warned that the Abbott test gave false negatives.

Meanwhile, the investigation of some 40 scientists and administrators already charged (see *Nature* 375, 526; 1995) is drawing to a close, with a decision on their trials expected soon. The investigating judge, Marie-Odile Bertella-Geffroy, had asked for the criminal investigations to be extended into a wider-ranging inquiry, but this was turned down last week by the public prosecutor's office.

Declan Butler

US research safety system 'in jeopardy'

[WASHINGTON] The effectiveness of institutional review boards (IRBs), the linchpin of the US system for monitoring the safety of human experimental subjects, is "in jeopardy", according to the Department of Health and Human Services (DHHS). A burgeoning workload has led to hurried, under-informed reviews with inadequate follow-up, its inspector general claims.

A four-volume report, released at a Congressional hearing last week, says changes in the research environment since IRBs were instituted 25 years ago have led to a system of boards that are incapable of fulfilling their role in the expanding world of US research.

"We are offering a very loud warning signal," said George Grob, the deputy inspector general for evaluation and inspections, who testified at the hearing on behalf of the inspector general. "We are saying major change needs to be made."

But Gary Ellis, director of the Office for Protection from Research Risks (OPRR) at the National Institutes of Health, said the risk of "catastrophic" failure of the current system was slight. He added: "One thing IRBs are not in is jeopardy. Let us set aside any sense of peril, danger, hazard or menace."

IRBs review the ethics of federally funded experiments involving human subjects, as well as studies on drugs and medical devices that must be approved by the Food and Drug Administration (FDA). The report claims their independence is threatened by an expectation that they should support their institution's interests — raising money and prestige by encouraging clinical research — as well as protect human subjects.

The report says IRBs offer their members little education on government ethics rules,



conduct minimal review of research that has already been approved, and review first-time proposals hastily. "They review too much, too quickly, with too little expertise," it says.

Among its recommendations are that IRBs should be relieved of perfunctory, procedural requirements that bog them down, and should include more non-scientific and non-institutional members.

At the root of the current predicament, it says, is the extra workload created by fundamental changes in the structure of research, including a proliferation of multi-site trials and an increasing number of research proposals — which will grow further if generous funding increases for the NIH continue.

Grob said the IRB system was built for a system that largely did not exist any more, comparing it to a shield that was "brittle, strained and... even cracked".

Christopher Shays (Republican, Connecticut), who chaired the hearing as head of the human resources subcommittee of the

House Government Reform and Oversight Committee, said the report, if anything, understates Congressional concerns that IRBs are "in jeopardy of being overwhelmed by the weight and complexity of their work".

For instance, he said the system "failed miserably" in a recent New York case in which healthy black and Latino boys, the younger brothers of delinquents, were given the now-banned drug fenfluramine to study their aggressive tendencies (see box). Yet even Shays, a leading critic of IRBs, did not suggest Congress should intervene by writing new laws. Many necessary reforms, he said, were within the power of IRB sponsors and administrators to implement at once.

Grob pointed to increased commercial sponsorship: "IRBs feel pressure to accommodate these sponsors who are looking for quick turnaround of their research and for whom time is money," he said.

Speaking on behalf of the Association of American Medical Colleges, Robert Levine, a professor of medicine at Yale University, called it "most unfortunate" that the tone of the report — which did not catalogue specific ethical lapses and specifically said it claims no widespread abuses of research subjects — "conflicts with its substance". It seemed to portray a system in crisis yet yielded no evidence of harm, Levine said, warning that adverse publicity would make it even harder to recruit high-quality people for demanding, largely unpaid IRB work.

Bert Spilker, the senior vice-president for scientific and regulatory affairs at the Pharmaceutical Research and Manufacturers of America, declared the IRB system "sound" and "working well". He urged the adoption of modest reforms to improve IRBs' efficiency but without new laws. He challenged a recommendation in the report that IRBs monitor clinical trials in progress, saying it was not necessary or appropriate because either drug companies audit the trials themselves or hire others to do this, or the FDA audits them.

Spilker said the notion that IRBs face conflict because of allegiance to their institutions was more theoretical than real. Levine said it was not in the interests of institutions to have their IRBs approving experiments that led to bad press or lawsuits.

Shays called it "pathetic" that the OPRR, which is charged with ensuring the protection of human subjects in all research conducted or funded by the DHHS, has only one full-time investigative worker.

Among its 70 current investigations, the OPRR carried out only one on-site investigation in the past year, Ellis said. That, at the University of Maryland, found that the IRB was understaffed, undereducated and has been meeting and approving experiments without a quorum.

Meredith Wadman

Review board head defends fenfluramine tests

[WASHINGTON] B. Timothy Walsh, a chair of the IRB that approved experiments in which fenfluramine, a now-banned diet drug, was given to healthy boys from ethnic minorities, defended his panel's decision last week before a Congressional hearing examining IRB effectiveness (see *Nature* 392, 747; 1998 & 393, 406; 1998.)

Walsh, who co-chaired the IRB at the New York Psychiatric Institute in Manhattan, told the human resources subcommittee of the House Government Reform and Oversight

Committee that his IRB concluded that a single, oral dose of fenfluramine presented risks that were "minor at most".

There was "no indication", he said, that a single dose of the drug causes heart valve damage like that seen in adults who took it for months, prompting its recall in 1997, two years after the New York experiment was stopped.

He added that the study of aggression risks in the boys, aged six to ten years old, was a type of research "critical for our country".

But Congressman

Edolphus Towns (Democrat, New York) called the IRB's approval of the experiment "very, very troubling". He produced an early proposal from the investigators in which white children were excluded; Walsh said the IRB required the investigators to change it to include all races.

Despite this, Towns complained, all of the study participants were black or Latino. "These children were chosen by design, not by chance," he said. "If this case is indicative, the IRB process... needs to be torn down and rebuilt from scratch, not reformed." **M.W.**

Job discrimination based on genetics set for California ban

[SAN FRANCISCO] The California senate is expected to finalize legislation this week to prohibit employers from discriminating on the basis of genetic characteristics. The bill, which the State Assembly approved last week, was expected to go to the governor Pete Wilson before Friday for a final signature.

The law would prevent employers from using any genetic information — including tests, family histories or conjectures from someone's appearance or ethnicity — in making hiring, promotion or other job-related decisions.

California already has one of the country's strongest laws prohibiting insurance-related genetic discrimination. State senator Patrick Johnston (Democrat, Sacramento/San Joaquin), who shepherded that legislation four years ago, has been pursuing restrictions on gene-related employment discrimination ever since.

"Job applicants should not be evaluated based on their genes," Johnston told the Assembly Appropriations Committee. "They should be evaluated based on their performance, their current health status, and their actual capacity to function in a job."

Johnston said several studies suggest that employers are using genetic information to deny employment to individuals. Companies may be attempting to avoid high health-insurance rates or a loss of productivity once a genetic disease strikes.

The Chamber of Commerce and the California Manufacturers Association opposed the legislation, but the pharmaceutical company SmithKline Beecham supported it. The Biotechnology Industry Organization has argued against a patchwork of non-discrimination laws at the state level, pointing out that there is already some federal protection, such as the Americans with Disabilities Act.

Johnston said that state laws on genetic discrimination are necessary because the existing act may not fully protect workers. In addition, the federal Disabilities Act covers only companies with at least 15 employees; the new law would apply to any employer of more than five people. The California legislation could lead the way for other states to enact strict laws against genetic discrimination in employment.

At the national level, a bill sponsored by Representative Louise Slaughter (Democrat, New York) and Senator Olympia Snowe (Republican, Maine) is scheduled for markup in the US Senate Labor and Human Resources Committee this week. Legislators are expected to vote on the bill, which is intended to prevent discrimination by health insurers on the basis of genetic information, within a month.

Sally Lehman

Israel reaches deal to join EU Framework programme

[JERUSALEM] Despite objections from some industrialists and government officials, Israeli Prime Minister Binyamin Netanyahu decided this week that Israel should participate in the European Union's fifth four-year Framework programme of research (FP5), which begins next year.

Netanyahu authorized the Israeli negotiating team to initial the agreement it had reached with the EU. But the agreement still requires the approval of the cabinet and of the finance committee of the Knesset, Israel's parliament.

The negotiating team, led by Orna Berry, chief scientist at the Ministry of Commerce and Industry, achieved what Berry said were significant improvements over the rules Israel worked under in the fourth Framework programme (FP4). This would put Israel on a par with the non-EU European states that participate, such as Norway.

Berry has been a strong supporter of Israel's continued participation, as a way of strengthening the country's presence in the European market for high-technology goods. But she does not dispute the claim levelled by the programme's opponents — that Israel is getting terms that are less good than those enjoyed by EU member states.

The FP5 rules require, for example, that Israel can only participate in projects involving at least two EU countries, while EU members need only find one other member state as a partner. But she argues that Israel cannot afford to remain outside the programme.

Israel's high-tech industries must understand, says Berry, that Framework is a European programme not designed with the interests of non-EU countries in mind. "We pay in advance, and we have to adjust ourselves to the EU's requirements," she says.

Berry points out that Israel has a trade deficit of US\$10 billion with Europe, half of the total value of the country's exports to the EU. She says Israel has already, through its membership of FP4, invested money and effort in becoming familiar with the programme and the Brussels bureaucracy. Not joining FP5 would mean writing off that investment, she says.

Berry's arguments are supported by many in universities. "Israel must join [FP5]," says Zehev Tadmor, president of the Technion — Israel Institute of Technology. "Europe is our natural partner and it would be absurd if we were not in the programme."

Technion researchers are involved in 17 research projects funded under FP4, in fields as diverse as aeronautics, medical technology and water desalination. "True, it's a big bureaucracy, but we've got to overcome that," says Tadmor.



Netanyahu: backs Framework deal despite claims that the rules discriminate against Israel.

But the programme's critics in industry and government disagree. Officials at the science and finance ministries have been arguing that the money Israel will have to invest in the programme — about \$45 million annually over four years — could be better spent in direct grants to local research projects. Some leading figures in the country's high-tech industries have been overtly hostile.

Zohar Zisapel, chairman of Rad Data Communications and of the Electronic Industries Association, argues that "Israel has been accepted into the programme under discriminatory conditions, and as a result has not been able fully to realize the opportunities it presented".

Zisapel points out that, under the rules covering the participation of non-EU states in FP4, Israeli companies have not been able to act as the coordinators of research programmes. Such restrictions come on top of the inherent disadvantages Israel faces as a non-EU member of the programme.

He suggests that Israel should be given not just a level playing field, but one tilted in its favour, arguing that European high-tech companies "have a lot more good reasons to want to cooperate with Israeli companies" than Israeli companies have to want to cooperate with European ones.

Uri Shor, spokesman of the Ministry of Science, says Israel should participate in the programme, but only on the same conditions enjoyed by member states. "We won't invest the money of Israeli taxpayers for other people's research programmes."

The prime minister's science adviser, Israel Hanukoglu, says: "We pay the full assessment but we don't get the whole programme. We're very interested in participating, but we want equal conditions."

However, the programme's opponents are unlikely to campaign against the agreement's ratification by the cabinet and Knesset.

Haim Watzman

AP/MATI STEIN

State department to hire science adviser

[WASHINGTON] The US Department of State is to appoint a science adviser in response to widespread criticism that it has been neglecting science and technology. Officials at the department, which is responsible for US foreign affairs, say the individual appointed will be a "distinguished person" from the science community.

Melinda Kimble, assistant secretary of state for oceans, environment and science, told a meeting at the National Academy of Sciences in Washington last week that the adviser would work directly with senior officials at the department, including herself and Stuart Eisenstadt, the under-secretary for global affairs, to "work with scientific issues at the state department and to develop a science policy" for it.

Kimble said after the meeting that the recruitment of the adviser was at a preliminary stage, but that she hoped to select someone this summer.

The meeting was held to introduce a new science and technology agreement between the United States and the European Union (see *Nature* 390, 543; 1997). Kimble told the meeting that problems with the department's handling of science issues had arisen because of overall cuts in its budget.

She claimed that the budget had fallen by \$5 billion since 1985, "when science and technology seemingly were a higher priority on our national agenda". (According to the 1999 budget book, the department's budget



Kimble: budget cuts caused problems for the state department's handling of science issues.

in fact expanded from \$2 billion in 1985 to \$5 billion last year, approximately in line with inflation.)

Kimble said there were "certainly some grounds" for recently expressed concerns and criticism "due to the perception that we are de-emphasizing science". She added that the department had responded to pressure on its resources by concentrating on "urgent" missions, but that it was now trying harder to balance these against longer-term concerns, such as science and technology.

The department has already asked the

national academy for advice on how it can better integrate science and technology into its sprawling operations (see *Nature* 392, 427; 1998). Glenn Schweitzer of the academy's international office says that, although a committee is yet to be appointed, its study will produce a short report by September to feed into next year's budget process. Its main report will follow in 1999.

Schweitzer told a meeting of the President's Council of Advisors on Science and Technology (PCAST) last week that, after 15 meetings with state department officials, he believed that there was "genuine interest at the political level and some interest among the career staff" in the outcome of the study.

Wendy Sherman, a counsellor in the office of Madeline Albright, the secretary of state, was unenthusiastic about the suggestion from one PCAST member that the state department could learn something from its counterparts abroad.

"Other countries aren't expected to carry the same range of responsibilities," she said. "It is probably easier for some of them to put more effort into science and technology because they don't have to cover the entire waterfront."

At a congressional hearing in March, it was suggested that the state department should take a leaf out of other nations' books by hiring science attachés on secondment from science agencies, rather than relying on career diplomats.

Colin Macilwain

Senate's proposed budget increases disappoint NSF and NASA

[WASHINGTON] The US Senate Appropriations Committee last week approved modest budget increases in the next financial year, which starts on 1 October, for the National Science Foundation (NSF) and the NASA space agency. But, for both agencies, this first step in the annual congressional budget process was a minor disappointment.

For example, although the senators proposed giving \$3.64 billion to the NSF next year, a 6.3 per cent increase on this year, it is \$129 million less than had been requested by the Clinton administration.

The White House had proposed a 12 per cent increase for the research account that funds individual grants to scientists, but the Senate committee recommends giving \$2.7 billion, an increase of only 7 per cent. Included in that are more than \$50 million in targeted 'earmarks', which would further dilute spending on basic research grants.

The earmarked funding includes an additional \$24 million for logistics support of Arctic research, whereas the NSF had asked for only \$9.5 million. The senators would also give \$50 million to research into

plant genomes, \$10 million more than the agency requested.

Two important projects, the Large Hadron Collider and the Millimeter Array telescope, would be fully funded. But the Senate committee denied a \$25 million request for the Polar Cap Observatory, which has been turned down for two years running by Senator Ted Stevens (Republican, Alaska), chairman of the appropriations committee. Stevens wants the observatory built on US territory, not in Canada as the NSF recommends.

The agency was also denied a 5 per cent increase for salaries and expenses, most of which was aimed at converting the 'FastLane' electronic grants administration system to full operational status.

The Senate committee recommended boosting NASA's budget request by \$150 million to \$13.6 billion. Space science would receive \$50 million of this increase, including a \$20 million boost for the Mars exploration programme. Earth sciences would receive \$25 million more than NASA asked for, and life and microgravity sciences

would get the amount requested.

Perhaps more important than the actual increases, the senators strongly warned NASA against "raiding" other parts of its budget to bail out the financially troubled space station. They also restructured the appropriations accounts in part "to ensure that Congress and this subcommittee gets honest figures for the [station]," according to Senator Christopher Bond (Republican, Missouri), who heads the NASA-NSF appropriations panel.

Funding for the space station, space science and aeronautics would now have separate congressional budgets. But NASA is expected to protest against this restructuring.

Both agencies will have further opportunities to plead their case. The full Senate is expected to vote on the appropriations bill later this month. Appropriations committee members in the House of Representatives will write their own funding bill this week, and the two versions will be reconciled into a final bill in the autumn.

Tony Reichhardt

Japanese media fuel fears of 'endocrine disrupters'

[TOKYO] Scientists in Japan claim the public is overreacting to concerns about the possible effects of 'endocrine disrupters', synthetic compounds suspected of disrupting human reproductive functions. They argue that media reports describing them as "deadly" or "fatal", despite the lack of clear scientific evidence, are exaggerated and misleading.

Research into the effects of the chemicals, known in Japan as 'environmental hormones', has recently expanded substantially, largely in response to public concerns.

One source of concern was a report released in March by a Teikyo University researcher claiming that only one of 34 sperm samples obtained from healthy university students had a normal sperm count. The report, which implicated endocrine disrupters as a possible cause, has been criticized as unreliable, not least because it involved only 34 samples as opposed to 15,000 in a similar study in Scandinavia in 1992.

A subsequent report released by the National Health Institute, which concluded that endocrine disrupters could be extracted from styrene dimers and trimers, which are used to make instant noodle containers, triggered media reports claiming that eating 'instant food' can cause infertility.

Over ¥18 billion (\$125 million) has been allocated from the government's supplementary budget for research into endocrine disrupters (see *Nature* 392, 748; 1998), and various ministries and agencies plan to monitor harmful chemicals in the environment.

Last month, the Environment Agency announced plans for a framework for tackling the problem of endocrine disrupters, which are unregulated in Japan. Although 67 chemicals, including dioxins, bisphenol and polychlorinated biphenyl compounds (PCBs), are classified as endocrine disrupters, their effects on the human reproductive system remain unknown, so regulations cannot be imposed.

According to the agency, which received ¥7.3 billion for research into endocrine disrupters for this fiscal year, beginning in April, designing a framework will be the first step towards new regulations, which will be set when chemicals are confirmed to be harmful. Research into their effects will be carried out at a laboratory to be built this year at the agency's National Institute for Environmental Studies (NIES).

The Ministry of Trade and Industry and the Ministry of Health and Welfare, each allocated ¥1.5–1.65 billion for the research, will collaborate to develop a high-speed screening device to analyse and classify chemicals.

Other research will look into methods for calculating levels of endocrine disrupters in

"... our priority should be to take a balanced view and uncover the potential effect of dioxins on human developmental mechanisms"

Taisen Iguchi, Yokohama City University

the environment, and set measures to minimize their spread. Dioxins, for example, are released into the air by burning waste, and PCBs, found in electrical equipment, are extracted from rivers and stream sediments.

Scientists have launched an academic society to promote an interdisciplinary approach to the problem. At its first meeting, held in Tokyo last week, its chairman, Tsuguyoshi Suzuki, ex-president of NIES, explained that the society's main aim is to "help protect public health by providing expertise to environmental policy makers, and this can only be done through a multi-disciplinary approach".

Researchers in Japan have been emphasizing the importance of work on endocrine disrupters for years, but little attention was given to the problem until this year. "While it is encouraging to see active interest in the issue by the government and the scientific community, it is imperative that more basic research is done," says Chisato Mori, associate professor in anatomy at Kyoto University.

Many point out that most of the budget is allocated to constructing buildings and buying equipment, with very little for research itself. "The government's current effort must not end as a temporary measure," says Mori.

Others have expressed concern over what they claim is misleading media coverage. In extreme cases, the chemicals have been blamed for a recent increase in violent crime among Japanese youths.

Taisen Iguchi, a professor in functional physiology at Yokohama City University, accepts that the media helped bring the subject to light, but says it must also be blamed for creating public misunderstanding.

"Since dioxins, classified as 'endocrine disrupters', have long been a problem in Japan for their carcinogenic properties, people have concluded that other chemicals are equally dangerous," says Iguchi. "Our top priority should be to take a balanced, objective view of the subject and, most importantly, uncover the potential effect of dioxins on human developmental mechanisms."

Asako Saegusa

Canada putting its faith in consolidation in health sector

[MONTREAL] The Medical Research Council of Canada (MRC) is planning to combine the country's 16 main academic health science centres and all other academic elements of the healthcare system into an integrated body, the Canadian Institutes of Health Research.

At present, the various components are only loosely coordinated. The proposed body would bring together the 16 health science centres, 50 teaching hospitals and 65 research institutes, which together employ more than 150,000 people full-time. Their combined annual operating budget is about Can\$16 billion (US\$10.9 billion).

The draft of the plan says that creation of the institutes "would energize the health research enterprise the way Medicare did the hospitals and health professionals more than 30 years ago with linkages and coordination around agreed national principles, supported by internationally competitive funding levels."

It adds that the Canadian health research enterprise constitutes an enormous asset base with the potential to become an even more effective foundation for the healthcare system. Modern information technology allows a degree of collaboration and sharing "that would encourage great research synergies".

The new integrated network "would identify and set priorities among emerging health issues," says the draft. Information flow would be stimulated between basic and clinical research and practice. This would decrease the lag "between precept and practice, and expedite the translation of new knowledge into effective patient care."

Salary support would be provided to those engaged in important and innovative research, and research teams identified as capable of internationally competitive work would receive "internationally commensurate levels of funding".

The draft plan notes that health research funding in Canada has not kept pace with that in the other leading industrialized (G7) countries. Although this year's federal budget provided an increase in funding in MRC's base budget, the present level (Can\$267 million) is at 1994–95 levels.

Canada's healthcare system is operated by provincial and territorial governments, with about 25 per cent of funding coming from federal government tax transfers, which have been falling in recent years, and through direct cash payments. Canadian health spending, forecast at Can\$76.6 billion in 1997, is approximately Can\$2,500 per person.

Federal health minister Allan Rock has already discussed the plan with the MRC. All representative groups will now be contacted and a national conference will be held in October.

David Spurgeon

Europeans adapt to compete with US neuroscience body

Alison Abbott

After a series of annual meetings that have failed to challenge the attractiveness of those organized by its US rival, the European Neuroscience Association is changing its name – to the Federation of European Neuroscience Societies – and structure, and has made financial savings. There are already signs that these changes, to be formally announced at its meeting in Berlin later this month, will reverse the trend.

[MUNICH] When the European Neuroscience Association (ENA) meets in Berlin at the end of this month, it may prove that the traditional weakness and lassitude of European-level scientific societies is a curable disease.

The Berlin meeting is the first to benefit from radical changes introduced in the past couple of years to combat this malaise, caused at least in part by tensions between the ENA and national societies.

In future, for example, the ENA and the national associations have agreed to hold their main meetings on alternate years, thus avoiding the inevitable clashes caused by each holding separate annual meetings.

The ENA has also bent over backwards to allay fears of a takeover, and is turning itself into a fully fledged federation of national organizations, with individual membership fees being replaced by national society fees.

"We promised that the executive of the federation would simply act as civil servants, carrying out the will of the national societies," says Wolf Singer, a director of the Max Planck Institute for Brain Research in Frankfurt and a former president of the ENA, who was one of the main engineers of the changes.

New name

It was under Singer's presidency that demands for changes within the ENA reached a climax at its 1992 annual meeting. Despite being held in Munich, it was able to attract only 1,700 delegates. Far more European neuroscientists make the annual pilgrimage across the Atlantic to the autumn meeting of the US Society for Neuroscience, where registration fees are low, scientific programmes are strong, and attendances regularly top 12,000.

At first there was resistance from well-established national societies to the changes that others considered necessary to make the European meetings as attractive as their US counterparts. But Singer says the ENA has been "very sensitive" to their concerns that a stronger European society could try to dominate them.

Both a new name — the Federation of European Neuroscience Societies (FENS) — and a new structure will be announced formally in Berlin. The number of registrants alone suggests that the patient is on the way to recovery.

With the numbers registered already exceeding 3,500, and at least 4,000 delegates expected, the meeting will be twice as big as it has ever been. This is thanks mostly to a halving of the registration fee, and a strong scientific programme with more symposia and satellite symposia than has been the case previously.

The organization to be unveiled in Berlin will have two councils, one made up of the presidents of each national society, the other of individual members elected by a voting system that is weighted by the population of neuroscientists in each country.

The national societies were reluctant to shift from annual to biannual meetings.

They were concerned, for example, that the continuity of the important social role played by such meetings would be lost. But they have accepted it as a price of more European-level interaction.

Another change is that the European-level meetings will now be held in June rather than September to avoid clashing with US neuroscience meetings. In years when no meeting is held, FENS will organize specialist schools and workshops.

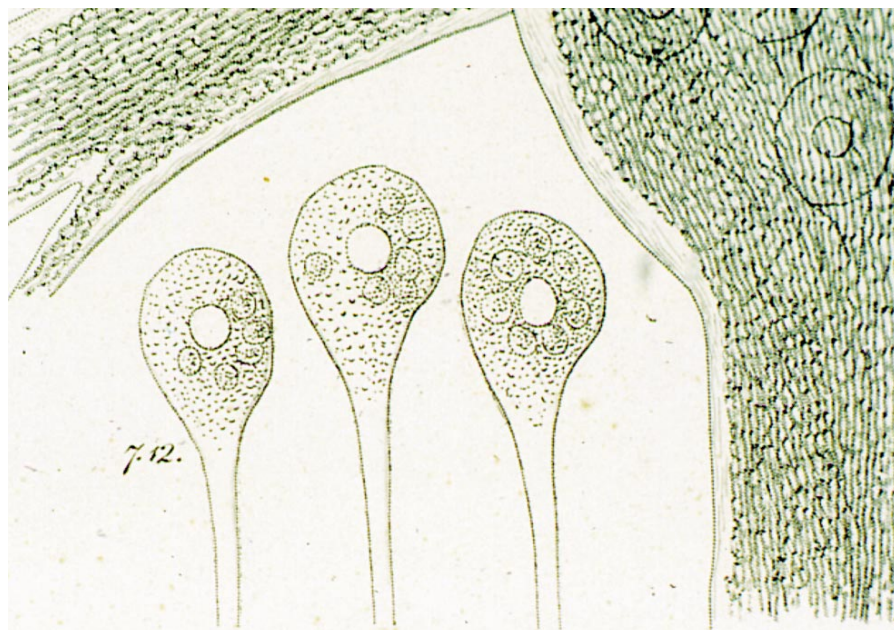
Anders Bjorkland, professor of neuroscience at the University of Lund and the present president of ENA, says this will help to compensate for the loss of the European Science Foundation's neuroscience programme, which was closed down a few years ago.

To help keep registration fees low, the ENA has abandoned the use of expensive professional conference organizers, and is now relying more heavily on the energy of local organizers. These will have the support of a new, and as yet very small, permanent secretariat in the Netherlands.

Money saving

ENA has saved some money by transferring its *European Journal of Neuroscience* from Oxford University Press to Blackwell, which will allow the association to keep a higher proportion of the journal's income.

Helmut Kettenmann, head of the department of cellular neuroscience at the Max Delbrück Centre for Molecular Medicine in Berlin, and organizer of the Berlin meeting on behalf of the German Neuroscience Association, says that the administrative burden of organizing the conference without professional help has been high. "But it is paying off," he adds.



First night nerves: these single neurons from the leech nervous system represent the first illustration of a neuron, dating from 1836. The image, from a book by C. G. Ehrenberg, is part of an exhibition about the history of neuroscience that runs alongside the ENA meeting at Charité hospital in Berlin.

Many journals have been persuaded to hold editorial board meetings in Berlin, as have many European neuroscience research networks, and several grant-giving organizations will have stands at the exhibition.

These factors alone could be sufficient to turn a traditionally dull ENA meeting into a lively event which organizers hope could also serve as an employment exchange. But Berlin also has an additional cultural dimension.

Support for the meeting

A national lottery grant to promote neuroscience locally during the meeting will pay for a series of popular films with neuroscience themes to be shown in local cinemas, each followed by discussions with scientists. This grant is also paying for an exhibition of contemporary artists — including a couple of Picasso and Klee originals — at a prestigious gallery in Potsdamer Platz.

The Institute for Anatomy in east Berlin's big teaching hospital, the Charité, is holding an exhibition about the history of neuroscience (see opposite). A grant from the Dana Alliance for Brain Initiatives — which funds the press work around the US neuroscience meeting — will pay for professional press services, another first for the ENA.

All the signs are that the meeting will be a success. But many neuroscientists are still awaiting the long-term outcome of the ENA's relaunch before accepting that a European society can be sufficiently powerful to compete effectively with the pull of the United States.

Indeed, according to Norman Bowery, professor of neuroscience at the University of Birmingham, the research community sees the Berlin meeting as "a sort of make-or-break for the European society".

Richard Morris, professor of neuroscience at the University of Edinburgh, points out that on average 3,000 British neuroscientists go to the US meeting every year compared with the 500 who attend ENA meetings, and this could be a hard trend to reverse.

But he finds the Berlin meeting very attractive, adding that sometimes the US meeting can be simply too big. "Four thousand attendees is the right number to allow you to be able to talk to your fellow scientists instead of just being able to wave to them across an aircraft hangar," he says.

The national societies, despite their early scepticism, are now fully behind the idea of FENS. The French society, for example, is providing 50 travel grants to allow its young members to attend.

Paul Bolam, professor of neuropharmacology at the University of Oxford and secretary of the British Neuroscience Association — one of the largest national neuroscience societies in Europe — says he is confident that negative feelings towards the ENA will be reversed by the Berlin meeting. □

Cell biologists set out on the path of reform

[MUNICH] The increasing economic and political integration of Europe has made many European-level scientific societies keen to increase their appeal to European scientists, who tend to view their national societies — followed by US societies — as their spiritual homes.

Particularly radical changes are now afoot for cell biologists in Europe, who are currently associated, through their national bodies, to a federation called the European Cell Biology Organization (ECBO).

According to Jean Grünberg, a professor of biochemistry at the University of Geneva, ECBO has "little power or flexibility" and does not have "a true European dimension". Most importantly, he adds, it does not reflect the full scope of cell biology.

A new society, called the European Life Sciences Organization (ELSO), will be launched in autumn, and will hold its first meeting in Geneva in 2000. It will include all molecular sciences relevant to cell biology, from developmental biology to neurobiology.

Although it has wide and enthusiastic backing within the community, ELSO is largely the brainchild of Kai Simons, senior scientist at the European Molecular Biology Laboratory in Heidelberg and director of the Max Planck Society's new Institute of Molecular Cell Biology and Genetics, currently being constructed in Dresden. But it has very broad support.

Simons has been elected as founding president of the society in a ballot organized by an ad hoc selection committee. He feels it is "time that scientists in Europe no longer have to go to the United States to attend a big meeting".

ECBO meetings, he says, have tended to suffer the malaise common to so many meetings of European-level scientific society meetings, namely low attendance, high registration fees and a variable quality of scientific programmes (see main article).

Simons — who himself comes from Finland — says that large, relatively inexpensive and high-quality European meetings are particularly important for young people from small countries who cannot afford to go to the United States, and who might not be invited to small specialist meetings.

To keep fees down, ELSO meetings will, like those of the recently revamped European Neuroscience Association (ENA — see opposite), avoid using professional conference organizers. To ensure that experience in organizing conferences is not lost, ELSO meetings will be rotated around three centres, one of which will be Geneva.

Like the ENA, ELSO is keen to offer strong scientific programmes and lively meetings that scientists cannot afford to miss. But, in direct contrast to the ENA (which is about to relaunch itself as the Federation of European Neuroscience Societies), it will do this by setting itself up as a society of individual members, given that ECBO already exists as a federation of national societies.

ELSO will model itself on the American Society for Cell Biology (ASCB), which covers all molecular sciences relevant to cell biology, and whose meetings can attract up to 10,000 researchers. Like the ASCB, it wants to wield influence as a lobbying force as well as to hold scientifically important meetings. These, it hopes,

will help young scientists by attracting enough people at both senior and junior level to act as a form of employment exchange.

Paul Nurse, director of the Imperial Cancer Research Foundation in London, is one of many who agree that that Europe's cell biologists, particularly young researchers, "need a Mecca equivalent to the ASCB, which does not require them to travel quite so far".

Another is Daniel Louvard, director of the research division of the Institut Curie in Paris and current president of ECBO. ELSO will not replace ECBO, he says. "But if it is successful, after a few years people are going to ask if we really need to have both — and FEBS [Federation of European Biochemical Societies] as well."

Other European scientific societies have also tried to launch, or relaunch, themselves as an important focus for the interests of their members, although so far with less success than the societies for cell biology and neurosciences.

The two European geological societies — the European Union of Geosciences and the European Geophysical Society — have been making efforts to fuse into a single body with the scientific and political strength of the American Geophysical Union. These have so far failed, apparently because they could not agree on its final structure.

But the European Federation of Pharmacological Societies (EPHAR), set up only a few years ago, could be at the bottom of the steep learning curve for running a European society. The success of its second meeting, to be held in Hungary next year, is being viewed as a touchstone for its survival.

A.A.

European countries set to agree on greenhouse targets

[LONDON] Members of the European Union (EU) will this week announce their individual targets to reduce greenhouse gas emissions under their obligations to the Kyoto protocol on climate change. Germany will cut its emissions the most — 22.5 per cent from 1990 levels before 2012. Denmark will make a 22 per cent reduction, with Austria making a 21 per cent cut by 2010.

The United Kingdom will reduce emissions by 12 per cent before 2012. France and Finland will be allowed to stabilize emissions at 1990 levels. In contrast, Portugal will be allowed to increase emissions by 24 per cent from 1990 levels, Greece by 23 per cent, Spain by 15 per cent and Luxembourg by 30 per cent.

These targets collectively are in line with the EU member states' joint commitment at last December's Kyoto climate conference to reduce emissions of a basket of six greenhouse gases by 8 per cent from 1990 levels before 2012.

Scientists seek a role in Ireland's peace deal

[DUBLIN] The Irish Research Scientists' Association has called on Ireland's research minister Bertie Ahern to propose an All-Ireland Science Council as one of the cross-border bodies to be created under the Northern Ireland peace initiative. In a letter last week to the minister, the association says that such a council would be a "potent vehicle to tap the benefits of science and technology for the good of all communities across the island". The association represents scientists from both sides of the border.

The association also asks the minister to use his influence to help persuade the Queen's University of Belfast not to close its geology department, as it has planned following its low ranking in the last UK research assessment exercise. The association argues that science departments should be given time to rebuild themselves in a peaceful society, where the environment will be more attractive to top scientists.

Appeal to US Senate to back test ban treaty

[WASHINGTON] Madeline Albright, the US Secretary of State, has called on the Senate to ratify the Comprehensive Test Ban Treaty. "Do not stall, do not delay, approve the CTBT," she declared during a major speech last week on disarmament.

Albright called on Senator Jesse Helms (Republican, North Carolina), chair of the Senate Foreign Relations Committee, to

examine the issue "on its merits" and offered to testify before the committee at any time. The speech was the opening shot in the Clinton administration's much-delayed campaign to win ratification of the treaty (see *Nature* **392**, 855; 1998).

But the speech may have served only to harden battle lines. Afterwards, Trent Lott (Republican, Mississippi), the majority leader in the Senate, lambasted the CTBT as "an unverifiable treaty overtaken by events". Tom Daschle (Democrat, South Dakota) promised to try to force a vote on it.

First success for Japan-US astronomical survey



[LONDON] The Sloan Digital Sky Survey, a joint Japanese-American project to build a three-dimensional map of the cosmos, has achieved 'first light'. The survey's telescope, at Apache Point observatory in New Mexico, recorded images (above) of a portion of sky in the constellations Serpens and Ophiuchus on the night of 27 May.

The telescope contains a digital camera to create five-colour images of the sky. This is complemented by two spectrographs, which provide information about the chemical composition, and the distances, of astronomical sources. The survey will record data for five years, producing a catalogue of positions and brightnesses for more than 100 million stars, galaxies and quasars, according to the survey's scientific director, Bruce Margon.

Antimatter experiment passes flight test

[WASHINGTON] Developers of the Alpha Magnetic Spectrometer say they are happy with a recent test flight aboard the space shuttle, despite problems with a high-rate data link. The controversial experiment is scheduled to be placed on board the International Space Station in 2002 to begin a three-year search for antimatter particles (see *Nature* **391**, 732; 1998).

A failure in the shuttle's communications system meant that only about 15 per cent of the data from the test flight, intended primarily to verify the spectrometer's detector, could be transmitted to the ground in real time. But all the data were safely

recorded on tape, according to principal investigator Samuel Ting of the Massachusetts Institute of Technology.

His group plans to run calibration tests in ground-based accelerators this year, but foresees no need for a further shuttle flight before launching to the space station.

Mudslide prompts call for geological surveys

[LONDON] Geological investigations should be made compulsory for all civil engineering projects, according to the Paris-based European Federation of Geologists. The federation is calling on European Union member states to consider legislation to this effect following the recent tragedy at Campania, Italy, where a mudslide caused 250 deaths in the Sarno valley.

The mudslide was caused partly by the fact that buildings had been erected over a disused seventeenth-century system of drainage. The federation says that landslides and foundation failures can often be predicted and prevented by a proper understanding of regional geological history.

Kyoto climate accord 'will wreak havoc in US'

[WASHINGTON] James Sensenbrenner (Republican, Wisconsin), chair of the US House of Representatives Science Committee, last week levelled some of the harshest criticism yet against the international agreement on climate change signed last December in Kyoto, Japan. Speaking at a forum on energy efficiency in Washington, Sensenbrenner said the treaty is "so flawed... that it cannot be salvaged".

While claiming that compliance with the treaty would "wreak havoc on US businesses and families," the congressman also attacked the treaty's scientific underpinnings. "Because the science is unclear, I don't think we should risk harming the American economy, and by extension the American people, by ratifying the Kyoto treaty."

San Jose sees its name written in the heavens

[LONDON] The Lick observatory on Mount Hamilton, east of San Jose, California, has had an asteroid named after its home city in recognition of officials' support for the observatory by controlling the interfering glow of streetlights.

M. R. C. Greenwood, chancellor of the University of California, Santa Cruz, and Joseph Miller, the observatory's director, presented the city council with a plaque carrying a digital image of the asteroid together with a tribute.

The Paris-based International Astronomical Union has approved the name.

Avenues of discovery in bioprospecting

Sir — The experts consulted by Colin Macilwain¹ provided useful evidence that bioprospecting for natural products is not the most productive way of finding biologically active compounds, but they did not explain why this should be so.

We would suggest that the high expectations for bioprospecting were partly based on the erroneous belief that a very high proportion of natural products must have some biological activity. However, the Screening Hypothesis², which we proposed to explain natural product chemical diversity, argues that specific, potent biological activity is a rare property for a chemical to possess, irrespective of whether a human, microbe or plant has done the synthesis. This is primarily because potent activity requires a very high degree of 'fit' between the chemical and a target receptor.

The hypothesis proposes that, in organisms that make natural products, there will have been a selection of biochemical traits in natural product biosynthetic pathways which enhance the generation and retention of chemical diversity. Because potent, specific biological activity is an extremely rare property for any molecule to possess, much of this natural product diversity will possess no inherent biological activity. This model proposes that organisms making natural products have been conducting combinatorial biochemistry and have been screening for activity for hundreds of millions of years³ before humans adopted a similar strategy. However, most organisms will be screening for biological activity with no particular relevance to human physiology.

So it is predictable that an efficient method of generating chemical diversity (combinatorial chemistry) coupled to high-throughput screening based on a cell process of known importance to human physiology will be more efficient than bioprospecting when searching for drugs.

Many compounds can be screened rapidly even though most will be inactive.

The opportunities that do exist for bioprospecting possibly lie more with the discovery of genes coding for enzymes involved in natural product biosynthesis, many of which might be expected to have a broad substrate tolerance. The addition of such genes to organisms with an existing rich natural product diversity should generate even more chemical diversity, producing chemical structures that currently lie beyond the scope of combinatorial chemistry. This opportunity to build on the existing combinatorial biochemistry of organisms will only be realized if the genetic diversity underlying the synthesis of natural products is explored and understood. So the bioprospecting for chemicals might have been relatively unsuccessful but bioprospecting for genes to use in combinatorial biochemistry might be more successful.

Richard D. Firn

*Department of Biology,
University of York,
York, UK*

e-mail: drf1@york.ac.uk

Clive G. Jones

*Institute of Ecosystem Studies,
Millbrook, New York 12545, USA
e-mail: clivejones@ecostudies.org*

Sir — An approach to improving the efficiency of identifying pharmaceutical lead compounds from natural products is to analyse and use ethnomedical information to guide the selection of plants. One example is traditional Chinese medicine. The strengths of the approach lie in the fact that Chinese medicine has a long history, was built on a theoretical (rather than purely empirical) basis, and its wealth of clinical experience has been recorded in literature.

In the United States, three new drugs of Chinese-herb origin have recently been

developed. They include a derivative of the antimalarial drug qinghaosu, the Alzheimer's drug huperzine A⁴, and the anti-HIV therapeutic trichosanthin⁵. These drugs may not finally make their way to the US market. But their novel prototype structures and new mechanisms of action have provided scientists with unique inspiration in the search for therapeutic agents.

Despite the vast gap between Chinese and Western medicines, decades of research has revealed that there is a degree of correlation between traditional indications of herbs and their pharmacological effects as defined by Western biomedicine. For instance, many herbs for 'eliminating heat and toxins' in the terms of traditional Chinese medicine possess antibacterial, antiviral, antipyretic and anti-inflammatory actions. A major mode of pharmacological actions of 'tonics' is to stimulate nonspecific and specific immune functions; and herbs for 'activating blood circulation and removing stasis' can improve blood flow dynamics, exert anti-thrombosis effect, and improve microcirculation.

The knowledge of such a relationship, which has been documented comprehensively⁶, should prove useful in a natural product drug discovery programme. If a project is set up to identify anti-thrombosis agents, for instance, then herbs used for 'activating blood circulation and removing stasis' ought to be examined.

Tianhan Xue & Li Zhang

*154 Kingston Avenue, Suite E,
Santa Barbara,
California 93117, USA*

1. Macilwain, C. *Nature* **392**, 535 (1998).

2. Jones, C. G. & Firn, R. D. *Phil. Trans. Roy. Soc. B* **333**, 273 (1991).

3. Firn, R. D. & Jones, C. G. *Recent Advances in Phytochemistry* (eds Romeo, J. T. et al.) 295–312 (Plenum, 1996).

4. Borman, S. *Chem. & Eng. News* 20 Sept., p. 35 (1993).

5. Byer, V. S. et al. *Aids* **4**, 1189 (1990).

6. Chen, C. (ed.) *The methodology for investigation on pharmacology of Chinese herbal medicine* (People's Health Publisher, Beijing, 1993).

Smarter than you think

Sir — Robert W. Old may be overestimating the difficulties students may have with the book *Principles of Development* by me and others (*Nature* **392**, 325; 1998).

It is to be hoped that they will not read single sentences in isolation from the rest of the text. Just four lines up from the offending sentence referring to Toll receptor activation, we describe Spaetzle binding to Toll. Indeed, on this and on the previous page, and in several figures, it is made clear that Spaetzle is the ligand for the Toll

receptor. Students may be more intelligent than Old supposes.

Lewis Wolpert

*Department of Anatomy and Developmental
Biology, University College London Medical School,
Medawar Building,
Gower Street, London WC1E 6BT, UK*

Cancer controversy

Sir — Your story about Gina Kolata's cancer research article in *The New York Times* reports a malicious rumour for which there

is no evidence (*Nature* **393**, 104; 1998).

For the record, after reading *The New York Times* article, I called Ms Kolata (which is perfectly proper in my role as her literary agent), she submitted a two-page letter-proposal and then she withdrew it. The timing is such that she could not have been colluding with herself to hoodwink *The New York Times*. Her ethical reputation is among the highest.

John Brockman

*Brockman Inc., 5 East 59th Street,
New York, New York 10022, USA
e-mail: john@brockman.com*

No conference critique

Sir — As organizers of the fifth annual Euroconference on Apoptosis, held in Bingen/Rhine last autumn, we were surprised to read the article misrepresenting the quality and the costs of this Euroconference (*Nature* **392**, 211; 1998). If the article was meant to be advertising for a forthcoming conference organized by young postdocs on research in apoptosis, it may have taken a wrong twist.

The Euroconferences on Apoptosis, organized by the European Cell Death Organization, are part of a series of meetings supported by the European Union. The size of the conferences is restricted to around 120 participants. The costs of last year's meeting were within the range of similar conferences (such as the Gordon Conference or Keystone Symposia). However, 50% of the non-invited participants, including PhD students and young postdocs, received funding from the European Union.

Participants were selected based on submitted abstracts. Additional support for the invited speakers came from Germany's research council, the Deutsche Forschungsgemeinschaft. The three-day meeting had 17 invited speakers and 18 oral presentations selected from submitted abstracts. Two poster sessions were held. All invited speakers and most of the selected speakers came from leading laboratories in the field and the data presented were mostly unpublished and/or conceptually important.

According to the questionnaire filled in by participants, the atmosphere at the meeting was congenial and the response enthusiastic. No critique related to the article in *Nature* was brought to the attention of the organizers. We feel that this clarification is important to ensure that the meeting series is not jeopardized. We do, however, appreciate and support the initiative of young postdocs to facilitate exchange at the planned First European Workshop on Cell Death.

Klaus-Michael Debatin

Peter H. Krammer

University Children's Hospital,
89070 Ulm, Germany

Environmental costs of subsidies oversimplified

Sir — Norman Myers, in his recent Commentary article, makes some cogent but somewhat simplistic observations about the economic and environmental costs that subsidies can inflict on society

(*Nature* **392**, 327–328; 1998). I should like to add some comments.

First, a sharp distinction should be drawn between an explicit, market-distorting public subsidy (for example, to Indian electricity users or Californian irrigators) and the many 'externalities' that characterize economic activity (agricultural pesticide runoff or urban congestion). In the former case, the marketplace provides clear-cut signals of the corrective action needed, however unpopular such action may be politically. In the latter case, although there has been significant progress in 'internalizing' externalities (for example through allowance trading under US limits on sulphur dioxide emissions), putting a monetary value on many externalities remains inexact and controversial.

Second, the dominant dilemma in fisheries is open access — a 'global commons' problem requiring cooperative approaches not amenable to being viewed in a subsidy framework.

Third, environmental improvement associated with removal of subsidies in one place may be diminished by developments elsewhere; German cutbacks in its highly protected coal industry may induce increased coal production in other countries and imports from abroad.

Finally, the beneficial subsidies of which Myers approves, such as those for renewable energy, may, on closer examination, be not so beneficial. The United States provides a handsome subsidy to corn processors producing ethanol, a costly policy whose consequences for an improved environment remain quite unclear.

Joel Darmstadter

Resources for the Future,
1616 P St NW,
Washington DC 20036-1400, USA
e-mail: darmstad@rff.org

Congressional hearings on genetics research

Sir — You make the inference in your News item about the multicultural task force on genetics and ethnic minorities that the American Association for the Advancement of Science has an agreement with the group to organize legislative hearings (*Nature* **392**, 428; 1998).

Following the involvement of the association in the two-day symposium sponsored by Howard University and Sinai Hospital, the organizers did contact my programme to learn about our experience in sponsoring non-partisan, legislative issue briefings in Congress.

Although we support the concerns of the task force and its desire to convene a

briefing on this important topic, at no time was it discussed that the association would be one of the organizers.

Mark S. Frankel (Director)

Scientific Freedom, Responsibility and Law Program,
American Association for the Advancement of
Science, 1200 New York Avenue NW,
Washington, DC 20005, USA
e-mail: mfrankel@aaaas.org

An art form whose time has come

Sir — Martin Kemp's series on Art and Science has been very welcome and I commend *Nature* for bringing this issue to the attention of readers.

A couple of Kemp's articles have focused on artists who have reached outside the studio and used nature as their artistic medium; I am thinking in particular of "Turrell's tunnelling"¹. Perhaps the best-known work of 'Earth-art' was created in 1970 by the American artist Robert Smithson (1938–73). I am referring to his monumental *Spiral Jetty*, a counterclockwise coil of rock and earth, 1,500 feet long and approximately 15 feet wide in the Great Salt Lake, Utah.

The spiral form is, of course, common in the natural world, such as mollusc shells; but, more importantly, the *Spiral Jetty* raises issues of time and scale, central theses in Smithson's work. As Smithson wrote, "each cubic salt crystal echoes the *Spiral Jetty* in terms of the crystal's molecular lattice. Growth in a crystal advances around a dislocation point, in the manner of a screw. The *Spiral Jetty* could be considered one layer within the spiralling crystal lattice, magnified trillions of times"².

Smithson, like his contemporary Donald Judd (1928–94), had a strong interest in geology and especially the nature of geologic time.

The *Spiral Jetty*, now submerged, continues to evolve as it is eroded, until nothing will remain. As an ephemeral work of art it confronts time, and forces us to think about the duration, not only of art, but also of geological processes and the shaping of the geologic past.

Smithson sought an art form that could encapsulate the furthest reaches of time and space; a purist notion, indeed, but one most worthy.

John Dalton

Department of Geosciences,
University of Texas at Dallas,
Richardson, Texas 75083, USA
e-mail: dalton@utdallas.edu

1. Kemp, M. *Nature* **391**, 131 (1998).

2. Smithson, R. in *Robert Smithson: The collected writings* (ed. Flam, J.) 143 (Univ. California Press, Berkeley, 1996).

Putting it on plastic

Karl Ziemelis

The quest to make integrated circuits entirely out of plastic has reached a landmark: functioning wafers have now been made that contain many such circuits. Their main advantages over much faster silicon electronics are simple — all-polymer devices are cheap and bendy.

The use of organic materials in the fabrication of electronic circuitry has a long history, particularly in lithography and packaging. Often overlooked, however, is the fact that some organic materials can behave as functional electronic materials. Important examples are the 'conjugated' polymers — polymeric materials that, by virtue of their unusual bonding structure, have many of the electronic properties associated with inorganic semiconductors. New work from the Philips Research Laboratories at Eindhoven, Netherlands, has brought electronic applications of these exotic materials closer to realization. As they describe in *Applied Physics Letters*, Drury *et al.*¹ have fabricated all-polymer integrated circuits on a flexible wafer (Fig. 1),

using a remarkably simple process.

The electronic properties of conjugated polymers fall well below the expectations and requirements of most of the traditional microelectronics industry, where crystalline silicon of high purity and structural integrity dominates. So it is most unlikely that we will see a plastic revolution in, for example, microprocessor development. But new markets are opening up for which the performance criteria are very different, and it is here that the polymers may find a niche. In particular, the combination of potentially low cost and mechanical flexibility makes these materials ideal for disposable, high-volume, electronic products — such as 'smart' labelling of consumer items, or electronic watermarks on paper documents.

Progress in developing polymer devices has been rapid. By replacing the inorganic semiconductor in conventional device structures with a semiconducting polymer, the basic operation of these materials has already been demonstrated^{2,3}. And operating speed, which is determined by charge-carrier mobility, is also improving. Even though the mobilities achieved in these devices are far from spectacular when compared with crystalline silicon, values approaching those of amorphous silicon have been obtained^{4,5} — a noteworthy achievement.

To make full use of the mechanical advantages afforded by these materials, flexibility needs to be imparted to all elements of the device, such as the electrodes and dielectric insulator, not just the semiconductor layer. There has already been some success in producing flexible devices^{6,7}. But for real applications, individual devices do not suffice. Many devices need to be integrated into circuits, where they can drive one another and so perform logic operations, while retaining their mechanical flexibility; and the fabrication route needs to be both straightforward and potentially cheap to implement. The group at Philips¹ have achieved both aims.

The cross-section through a transistor in Fig. 2 shows how the wafer is built up. First, there is the substrate, a foil of an electrically inert polymer that supports the devices. On this the various material layers — lower electrical connections, semiconductor, dielectric insulator and upper electrical connections, all of which are polymeric — are cast in turn from solution, and patterned if required.

But that does not do full justice to the ingenuity underlying the new fabrication scheme. For a start, dissolution of the underlying layers must be avoided at each deposition step, if the integrity of the device structures is to be maintained, so the choice of polymers and solvents used for each layer is critical. Even more interesting is how the electrical connections are patterned. The polymer used is polyaniline, doped with camphorsulphonic acid to achieve levels of conductivity sufficient for electrical connections. The twist comes with the inclusion in these films of a photoinitiator molecule, which after ultraviolet irradiation increases the resistance of the polyaniline by more than ten orders of magnitude. So the complex pattern of conducting leads and electrodes required for an integrated circuit can be produced without removing unwanted conducting material: each polyaniline layer is simply exposed to ultraviolet radiation through an optical mask. To prevent further unwanted reactions with the unexposed, conducting parts of the film, the remaining photoinitiator can be removed through sublimation simply by heating the film. This leaves a surface that is still flat enough for deposition of subsequent layers.

Finally, to integrate the individual devices into a functional circuit one must add vertical interconnections — electrical contacts between the top electrode of one transistor and the bottom electrode of another. At face value, the solution adopted by the authors may seem crude, but it avoids the need

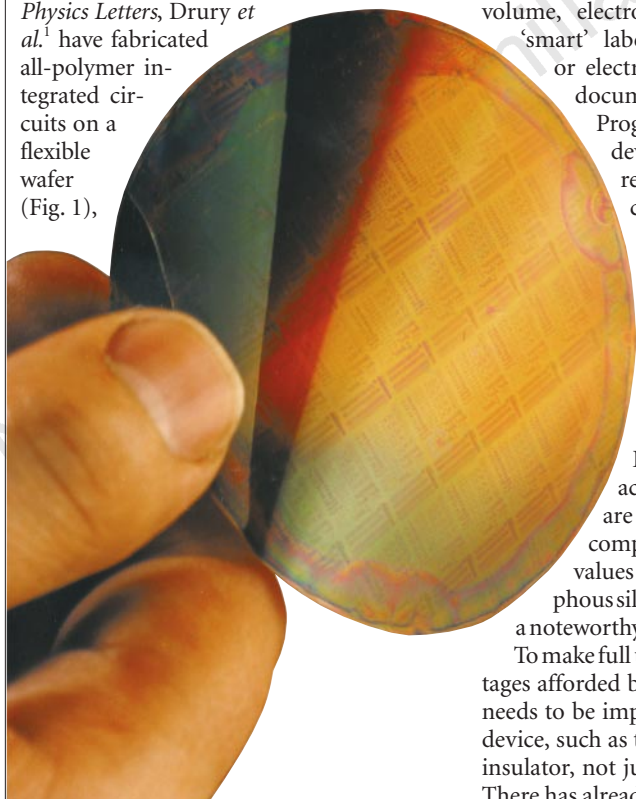


Figure 1 Flexible friend — a pattern of circuits and test structures is repeated about 50 times over the surface of this polymer wafer. The most complex is an integrated circuit (a 15-bit mechanically programmable code generator) containing 326 transistors and over 300 vertical interconnections, which remains fully operational even when the substrate is sharply bent.

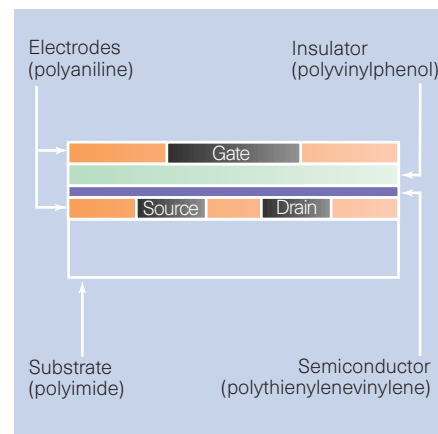


Figure 2 Slice through a plastic transistor. In the technique used by Drury *et al.*¹, layers of polymer are cast one by one on a spinning disk, and the electrodes are patterned by ultraviolet light.

for subsequent patterning of the film: they simply punch through the layers with pins (albeit in a controlled manner), which produces enough mixing between the two polyaniline layers to make usable vertical electrical contacts.

The simplicity of the fabrication process, compared with ordinary silicon lithography, promises very cheap circuits in the future. As a vivid illustration of the feasibility of the approach, Fig. 1 shows a 3-inch polyimide 'wafer' on which Drury *et al.*¹ have built a variety of simple test structures and more complex integrated circuits.

The levels of performance routinely obtained in the microelectronics industry are impressive, and it is unlikely, at least in the foreseeable future, that silicon will be displaced from its position as the semiconductor of choice. And certainly not by a polymer: by the usual measures, the performance of the all-polymer circuits built so far is embarrassingly low (an operating frequency of

only 30 bits per second, in the case of the most complex circuit patterned onto the wafer in Fig. 1). But computational speed and power are not essential for the applications that the proponents of plastic electronics have in mind. With relatively modest improvements in performance, smart(ish) electronic labels making use of polymeric circuitry could become as commonplace as the barcode. □

Karl Ziemelis is Physical Sciences Editor of Nature.
e-mail: k.ziemelis@nature.com

1. Drury, C. J., Mutsaers, C. M. J., Hart, C. M., Matters, M. & de Leeuw, D. M. *Appl. Phys. Lett.* (in the press).
2. Burroughes, J. H., Jones, C. A. & Friend, R. H. *Nature* **335**, 137–141 (1988).
3. Assadi, A., Svensson, C., Willander, M. & Inganäs, O. *Fuch Phys. Lett.* **53**, 195–197 (1988).
4. Fuchigami, H., Tsumura, A. & Koezuka, H. *Appl. Phys. Lett.* **63**, 1372–1374 (1993).
5. Gundlach, D. J., Yin, Y. Y., Jackson, T. N., Nelson, S. F. & Schlom, D. G. *IEEE Elec. Dev. Lett.* **18**, 87–89 (1997).
6. Gustafsson, G. *et al. Nature* **357**, 477–479 (1992).
7. Garnier, F., Hajlaoui, R., Yassar, A. & Srivastava, P. *Science* **265**, 1684–1686 (1994).

Circadian rhythms

New cogwheels in the clockworks

Ueli Schibler

Many organisms, from bacteria to man, contain self-sustained circadian oscillators that control daily fluctuations in behaviour and physiology. These oscillators are controlled by a number of clock genes, but the nature of the interplay between them is not known. Now, however, four papers — two by Rosbash and Hall and colleagues in *Cell*^{1,2}, and two by Weitz, Takahashi, Kay and co-workers in *Science*^{3,4} — address this issue. These landmark studies not only add a new, essential gene to the

circadian oscillator, but they also show how the cogwheels may be connected to make the clock tick in flies and mammals.

The first demonstration that circadian behaviour is driven by defined genes came in 1971, when Konopka and Benzer⁵ discovered a locus called *period* (*per*) in the fruitfly *Drosophila melanogaster*. Mutations in *per* lead to altered or abolished daily rhythmicity of locomotor activity and emergence from the pupa case. More than 20 years later, mutations in another locus, *timeless* (*tim*),

were found⁶ to result in similar disturbances in circadian timing. Then, last year, *Clock* was isolated⁷ — a murine gene whose function is essential for correct and persistent circadian timing. Moreover, three murine genes (*mper1*, *mper2* and *mper3*) were found^{8–10} to have striking sequence similarity to the *Drosophila per* gene.

The circadian clock is made up of three components — an input pathway that resets the clock, a central pacemaker that generates the oscillation, and an output pathway that translates the oscillations into overt rhythms in behaviour and physiology. Both flies and mammals contain master oscillators that coordinate the timing of circadian outputs such as locomotor activity (in *Drosophila*) or body temperature (in mammals). These master clocks have been localized in *Drosophila* to lateral pacemaker neurons in the head, and in mammals to the suprachiasmatic nucleus of the hypothalamus. However, many non-neuronal types of cell also contain self-sustained circadian clocks that continue to run in organ or tissue cultures^{11,12}.

Just how can a cell measure a 24-hour interval without receiving external time cues? The model, proposed by Rosbash and colleagues¹³ and extended by Young *et al.*¹⁴, involves a negative-feedback loop for the expression of *per* and *tim* (Fig. 1). Both *per* and *tim* messenger RNAs accumulate rhythmically, with a period of about 24 hours — identical to that of overt rhythms. The Per and Tim proteins form heterodimers that repress transcription of their own genes when these proteins reach a critical threshold concentration in the nucleus. Without transcription of the *per* and *tim* genes, levels of Per and Tim drop below the threshold, autorepression is

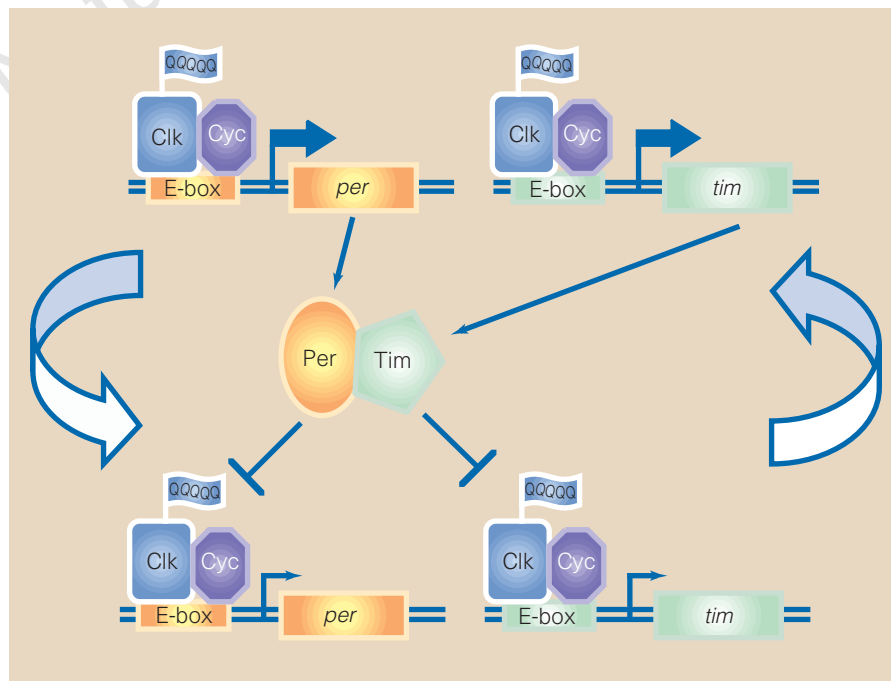


Figure 1 A negative-feedback model for cyclic expression of *per* and *tim*. The promoters of *per* and *tim* contain E-boxes with the core sequence 5'-CACGTG-3' that are bound by heterodimers of the transcriptional activators Clk (*Drosophila* Clock) and Cyc (*Drosophila* BMAL1). The *per* and *tim* messenger RNAs are synthesized until there is a high enough concentration of heterodimers formed by their protein products, Per and Tim, to repress *per* and *tim* transcription. Repression by the Per-Tim heterodimers acts through the E-box (or nearby sequences), presumably by inactivating the Clk-Cyc heterodimer. After Per and Tim have decayed, Clk and Cyc can again activate transcription, and a new circadian cycle of *per* and *tim* expression is initiated. Post-transcriptional regulation (for example, by phosphorylation of Per and Tim) may contribute to the activity and/or accumulation of these proteins. The glutamine (Q)-rich Clk activation domain is depicted as a flag.

relieved and transcription can start anew.

Obviously, such a negative-feedback loop can work only if the default state of the gene is 'active'. So what activates the transcription of *per* and *tim*? Last year, King *et al.*⁷ identified the mouse Clock protein, and found it to be a basic PAS helix-loop-helix transcription factor. (The name PAS comes from the proteins in which this domain was first identified.) The original mutation in the *Clock* gene eliminated a splice donor site, which resulted in the skipping (during splicing) of an exon that encodes part of a glutamine-rich transactivation domain⁷ (depicted as a flag with Q's in Fig. 1). The mutant Clock protein still contains its DNA-binding domain, but it cannot activate transcription efficiently. If the targets of Clock are essential oscillator genes such as *per* and *tim*, this mutation should result in altered or abolished rhythmicity of the affected animals, similar to that caused by mutations in *per* and *tim* themselves — and this is exactly what is observed for *Clock* mutant mice⁷.

Gekakis *et al.*³ have now used a yeast-based two-hybrid system to identify a functional dimerization partner for the mouse Clock protein, reasoning that basic helix-loop-helix transcription factors generally act as heterodimers. They identified a protein that matched three sequence entries in databases for a human PAS basic helix-loop-helix protein; BMAL1, JAP3 and MOP3. Pleasingly, the authors found that a Clock-BMAL1 heterodimer stimulates transcription from a *cis*-acting *Drosophila per* promoter element that is essential for circadian transcription. This DNA element contains an E-box, which is a known recognition sequence for basic helix-loop-helix proteins. The Clock-BMAL1 heterodimer also stimulates transcription from similar promoter sequences found in the 5' flanking sequences of the murine *per* homologue, *mper1*. As expected, the mutant Clock protein lacking part of the glutamine-rich activation domain failed to stimulate transcription.

Because the murine Clock-BMAL1 heterodimer can stimulate transcription through a *cis*-acting *Drosophila* DNA element, the authors predicted that *Drosophila* counterparts of these proteins may exist. Indeed, the corresponding *Drosophila* sequences — *dClock* (named *Clk* by the Rosbash laboratory) and *dbmal1* (dubbed *cyc*, for cycle) — were soon found in databases^{1,2,4}. Darlington *et al.*⁴ also identified *Clk* by screening a *Drosophila* complementary DNA library using mouse *Clock* as a probe. In co-transfection experiments with *Drosophila* S2 cells, these authors found that the *Clk* protein alone strongly activates transcription from the *per* promoter E-box, presumably because the level of endogenous *Cyc* protein is already high in these cultured cells. Importantly, ectopic co-expression of *Tim* and *Per* proteins results in a strong

attenuation of *Clk*-dependent transcription in transfected cells (for similar results in flies see ref. 15). Because basal expression of the *per* reporter gene — with or without a functional E-box — is not affected by *Per*-*Tim* heterodimers, repression probably occurs through inactivation of the *Clk*-*Cyc* complex (Fig. 1).

In transfected cells, effector molecules such as transcriptional activators or repressors often reach high, non-physiological levels. So, conclusions based solely on such assays have to be taken with a grain of salt. However, the genetic work reported by Allada *et al.*¹ and Rutila *et al.*² beautifully supports the predictions made by the model presented in Fig. 1. Several years ago, the Rosbash and Hall groups identified two loci in a recessive circadian mutant screen that mapped to positions 66A and 76C of the third *Drosophila* chromosome. These loci turned out to encode *Clk* and *Cyc*, respectively. The authors found that flies homozygous for either the *Clk* or *cyc* mutant alleles are arrhythmic and express very low and temporally constant levels of *Per* and *Tim*. Similar to murine mutant *Clock*, *Drosophila* mutant *Clk* encodes a protein that does not have a functional transactivation domain. Mutant *cyc* can be considered as a null allele.

This work on circadian pacemaker genes is spectacular not only for its heuristic value, but also because it illustrates the synergistic power of genetics and genomics (not for nothing that the first three letters are the same!). Given the evolutionary conservation

of the mechanisms that govern circadian timing, there is little doubt that database searches will continue to be successful in hunting down important pacemaker components across species. But there is much more to this than simply showing that "it's the same everywhere". *Drosophila* will probably continue to be the animal system of choice for genetics and, thus, for the isolation of new pacemaker genes. The biochemical dissection of the clock mechanisms, however, may be more readily accomplished in mammalian tissues or in cultured cells¹². □

Ueli Schibler is in the Department of Molecular Biology, University of Geneva, 30 Quai Ernest Ansermet, CH-1211 Geneva-4, Switzerland.
e-mail: Ueli.Schibler@molbio.unige.ch

1. Allada, R., White, N. E., So, W. V., Hall, J. J. & Rosbash, M. *Cell* **93**, 791–804 (1998).
2. Rutila, J. E. *et al.* *Cell* **93**, 805–814 (1998).
3. Gekakis, N. *et al.* *Science* **280**, 1564–1569 (1998).
4. Darlington, T. K. *et al.* *Science* **280**, 1599–1603 (1998).
5. Konopka, R. J. & Benzer, S. *Proc. Natl Acad. Sci. USA* **68**, 2112–2116 (1971).
6. Sehgal, A., Price, J. L., Man, B. & Young, M. W. *Science* **263**, 1603–1606 (1994).
7. King, D. P. *et al.* *Cell* **89**, 641–653 (1997).
8. Tei, H. *et al.* *Nature* **389**, 512–516 (1997).
9. Albrecht, U., Sun, Z. S., Eichele, G. & Lee, C. C. *Cell* **91**, 1055–1064 (1997).
10. Zylka, M. J., Shearman, L. P., Weaver, D. R. & Reppert, S. M. *Neuron* (in the press).
11. Plautz, J. D., Kaneko, M., Hall, J. C. & Kay, S. A. *Science* **278**, 1632–1635 (1997).
12. Balsalobre, A., Damiola, F. & Schibler, U. *Cell* **93**, 929–937 (1998).
13. Rosbash, M. *et al.* *Cold Spring Harb. Symp. Quant. Biol.* **61**, 265–278 (1996).
14. Young, M. W., Wager-Smith, K., Vossahl, L., Saez, L. & Myers, M. P. *Cold Spring Harb. Symp. Quant. Biol.* **61**, 279–284 (1996).
15. Zeng, H., Hardin, P. E. & Rosbash, M. *EMBO J.* **13**, 3590–3598 (1994).

Colloids

A surprisingly attractive couple

David G. Grier

A vast number of industrial and natural processes depend on controlling the interactions between micrometre-scale colloidal particles dispersed in fluid suspensions. Colloidal milk fat and proteins aggregate into either cheese or yoghurt depending on how they are handled; colloidal paint pigments must remain suspended for years in a can, yet should coagulate rapidly into a tough coating when spread on a wall. For just such reasons, colloidal interactions and their ramifications have been investigated intensively for two centuries.

And yet, over the past two decades, one of the fundamental tenets of colloid science has come under attack: contrary to the lessons of long experience, it now seems that like-charged colloidal particles sometimes attract one another. Reports of this extraordinary effect and efforts to explain it have sparked a rancorous debate. Its resolution may at last be at hand, in part because of results from a numerical study by Bowen and Sharif, reported on page 663 of this issue¹.

Although many processes affect colloidal behaviour, the phenomena under contention depend on just three energy scales — those fixed by van der Waals attraction, by the randomizing influence of thermal energy, and by the hierarchy of electrostatic interactions among highly charged colloidal particles and the singly charged simple ions around them. The balance of these three determines the properties of a large class of colloidal suspensions.

van der Waals attraction arises from sympathetic fluctuations in particles' charge distributions. It causes colloidal particles to aggregate, and is partly responsible for dust's tenacious grip on a television screen. But van der Waals attraction only exceeds thermal energies for colloidal particles very near to contact, at separations smaller than a few nanometres — more widely separated particles are saved from its grip by thermal collisions with surrounding fluid molecules (so preventing particles from colliding prevents them from aggregating). And

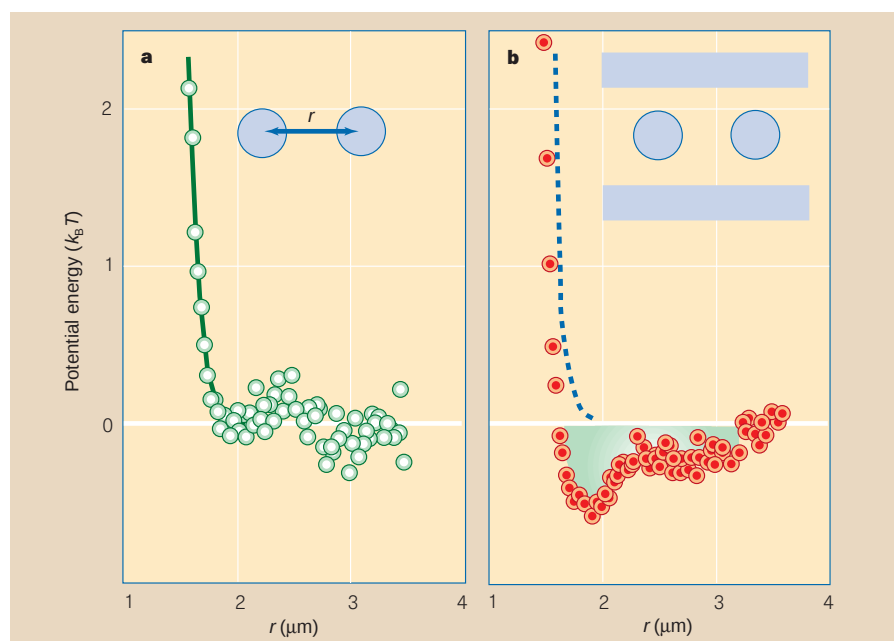


Figure 1 Attractions of confinement. The interaction potential between unconfined pairs of charged colloidal spheres is purely repulsive (a). Measurements⁶ (green circles) were made on two polystyrene sulphate spheres of radius $0.482\ \mu\text{m}$, and agree with the prediction of DLVO theory (solid line). But when the same pair of spheres is confined by parallel glass walls separated by $3.5\ \mu\text{m}$, an attractive minimum develops in the pair potential (b).

electrostatic forces can be much stronger: the conventional wisdom, worked out more than 50 years ago by Derjaguin, Landau, Verwey and Overbeek (DLVO), is that just a few static charges on each colloidal particle's surface can cause repulsions strong enough to keep the particles stably separated².

The system of equations describing interactions among the spheres, solvent and simple ions is so highly nonlinear that it resists analytical treatment to this day. The DLVO formulation avoids this complexity by linearizing the equations and averaging out ionic fluctuations. Despite these simplifying approximations, the DLVO theory accounts quite well for the stability and properties of charge-stabilized colloidal suspensions and serves as the standard model for colloidal electrostatic interactions.

The problem is that some highly charged colloidal particles do not behave as DLVO says they should. Inspired by the alternative Sogami–Ise theory for colloidal electrostatic interactions, Norio Ise and his collaborators have investigated the structure of colloidal suspensions. In observations spanning 15 years, they recorded a variety of phenomena apparently inconsistent with the DLVO theory, ranging from anomalously small lattice constants in colloidal crystals³ to large stable voids in colloidal fluids⁴. They interpreted their observations as evidence for long-range attractions between like-charged spheres. Other researchers suggested more conventional explanations, and controversy ensued.

Because many-body behaviour in suspensions often obscures underlying pair

interactions, measurements of bulk properties leave room for disagreement. Only in the past five years have techniques been developed that can measure the fantastically small forces between individual colloidal particles (Fig. 1). These measurements have begun to tell a coherent story: isolated pairs of like-charged spheres are found to repel each other much as predicted by the DLVO theory^{5–9}. But spheres confined by glass walls^{6,9,10} or by a concentration of other spheres⁹ (Fig. 2), develop long-range attractions inconsistent with DLVO. And attractions are favoured by

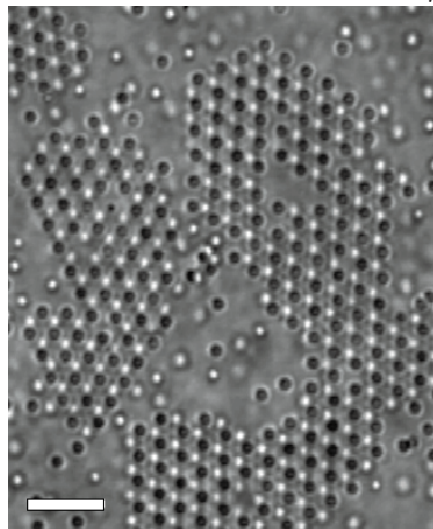


Figure 2 Superheated colloidal crystals (from ref. 9). Attractions similar to those in Fig. 1 are implicated in their formation, but here the confinement responsible for the many-sphere cohesion appears to be provided by the spheres themselves. Scale bar, $10\ \mu\text{m}$.

highly charged spheres in very low salt concentrations — circumstances under which the DLVO approximations might be expected to fail.

The apparent breakdown of linear superposition (pairs repel but groups cohere) implicates nonlinearity as the culprit. But a slew of alternative explanations have been proposed, including the inherently linear Sogami–Ise theory and mechanisms involving fluctuations in the simple-ion distributions. Differences between these theories hinge on the experimentally invisible simple ions.

Bowen and Sharif's numerical study¹ demonstrates that nonlinearity — long neglected — can indeed explain the observed attractions. The simple-ion distribution around an isolated pair of spheres mediates repulsive interactions, even in the nonlinear calculations. Confining walls, however, redistribute the simple ions so as to mediate a long-range attraction. (Under DLVO approximations, the comparable redistributions turn out to be too subtle to induce attraction.) So both nonlinearity and confinement are needed to yield attractions between spheres. Temporal fluctuations are not considered in these calculations, and therefore are not necessary.

Knowing the ingredients necessary for like-charge colloidal attractions to occur is a big step towards understanding their origin and predicting their ramifications. But we still do not understand how nonlinearity and geometric confinement conspire to produce simple-ion distributions conducive to long-range cohesion between neighbouring spheres. Numerical simulations, like physical experiments, offer insights into how particular systems behave under particular circumstances; they offer no predictive insights into the behaviour of other systems.

What geometries support long-range, like-charge colloidal interactions? How can we turn these interactions on and off? Could they affect smaller macroions such as proteins and DNA? By identifying the processes at work, Bowen and Sharif have brought us to the brink of being able to answer such questions. □

David G. Grier is at the James Franck Institute and Department of Physics, University of Chicago, 5640 South Ellis Avenue, Chicago, Illinois 60637, USA.
e-mail: grier@fafnir.uchicago.edu

1. Bowen, W. R. & Sharif, A. O. *Nature* **393**, 663–665 (1998).
2. Russel, W. B., Saville, D. A. & Schowalter, W. R. *Colloidal Dispersions* (Cambridge Univ. Press, 1989).
3. Yoshiyama, T., Sogami, I. & Ise, N. *Phys. Rev. Lett.* **53**, 2153–2156 (1984).
4. Yoshida, H., Ise, N. & Hashimoto, T. *J. Chem. Phys.* **103**, 10146–10151 (1995).
5. Crocker, J. C. & Grier, D. G. *Phys. Rev. Lett.* **73**, 352–355 (1994).
6. Crocker, J. C. & Grier, D. G. *Phys. Rev. Lett.* **77**, 1897–1900 (1996).
7. Vondermassen, K., Bongers, J., Mueller, A. & Versmold, H. *Langmuir* **10**, 1351–1353 (1994).
8. Sugimoto, T. *et al. Langmuir* **13**, 5528–5530 (1997).
9. Larsen, A. E. & Grier, D. G. *Nature* **385**, 230–233 (1997).
10. Kepler, G. M. & Fraden, S. *Phys. Rev. Lett.* **73**, 356–359 (1994).

Molecular motors

Keeping the beat

Alan J. Hunt

The undulating beat of eukaryotic flagella and cilia produces forces that propel cells (or the constituents of their environment) in creatures ranging from algae to humans. The beats arise from bends that travel from a flagellum's base to its tip — by repetitively generating these bends, flagella assume waveforms appropriate for their specialized functions. The timing mechanisms that coordinate these beats remain mysterious, but models indicate that the underlying metronome requires a complicated structure called the axoneme^{1–3}, at the core of the flagellum. On page 711 of this issue, however, Shingyoji *et al.*⁴ report that oscillatory movements can be generated by surprisingly simple sub-structures of the axoneme.

The axoneme is built from about 250 different polypeptide components, including nine long, thin doublet microtubules, arranged parallel to one another in a cylindrical array. Each doublet consists of an A-microtubule that is bound lengthwise to an incomplete B-microtubule (Fig. 1). Molecules of dynein, an ATP-fuelled motor protein, are arranged along the A-microtubule in two parallel rows that project towards an adjacent doublet. The doublets are locked together at the base of the axoneme so, when dynein generates shearing forces between them, they bend rather than slide apart.

Shingyoji *et al.*⁴ have used a focused laser beam (optical tweezers) as a micromanipula-

tor and force transducer. They find that the direction of the movement generated by a small number of dynein molecules (perhaps even one) on an isolated doublet microtubule prepared from sea-urchin flagella oscillates when the dynein molecules work against an elastic load. The frequency of the oscillations is similar to the beat of an intact flagellum, in contrast to the faster, smaller oscillations observed in intact non-beating axonemes⁵.

Understanding the geometry of the assay is crucial for interpreting these results. Shingyoji *et al.* used optical tweezers to manipulate a bead bound to a microtubule, so that this microtubule was positioned across a doublet microtubule at an obtuse angle (see Fig. 1a on page 711). Despite this unusual non-parallel orientation, the dyneins on the doublet were compliant enough to move the microtubule, displacing the bead within the optical tweezers. The effect of the optical tweezers was similar to that of attaching one end of the microtubule to a spring — the restoring force increased with the displacement.

How might these oscillations arise? If more than one molecule of dynein is involved, oscillations could result from variations in the number of motors that interact with the microtubule, as suggested in Fig. 1. As the motors displace the bead-bound microtubule, the restoring force from the optical tweezers will impart a torque on the doublet microtubule about its attachment to

the glass slide. Eventually, rotation of the doublet could cause a subset of the dynein motors to be pulled away from the microtubule. The remaining motors, unable to sustain their increased load, would reverse direction and 'walk' backwards along the microtubule until enough strain is released to allow the lost motors to re-engage — then, forward movement would resume.

This is plausible. Shingyoji *et al.* have found a decrease of speed in both directions with diminishing concentrations of ATP, indicating that dynein molecules can indeed walk backwards. This differs from the saw-tooth movement traces that are seen when an elastic load is placed on single molecules of kinesin, a motor that detaches from microtubules when overloaded^{6,7}. *Chlamydomonas* mutants that lack subgroups of dynein arms⁸ would be well suited for testing this model and for sorting out the biophysical contributions of different dyneins within an axoneme.

The most intriguing suggestion made by Shingyoji *et al.* is that the oscillations might be generated by a single molecule of dynein. If correct, this implies that a doublet-bound dynein molecule can retain a 'memory' of previous events, because it varies in a time-dependent manner in its response to an opposing force. Considering the structure of dynein, this is not a trivial task. Sea-urchin flagellar dyneins are large, multi-subunit molecules that consist of a stalk and two globular heads. The heads contain sites for microtubule binding and ATP hydrolysis. Because the heads are only about 13 nm in diameter^{9,10}, dynein probably takes more than one ATP-hydrolysing 'step' to span 32 nm, which is the average amplitude of the oscillations. So it is difficult to imagine how the oscillations might be simply associated with the different chemical states of dynein's chemo-mechanical ATP hydrolysis cycle.

Backward movements of a single dynein might persist until some unspecified strain is released from the dynein molecule. However, a simple elastic relaxation cannot adequately describe the data. Any strain that is large enough to be distinguished from random thermal fluctuations would be dissipated in microseconds at most, yet the period of the oscillations was many milliseconds. But perhaps there is a way to slow dynein's relaxation. For example, at peak strain, rotation of the doublet might pull a single dynein head away from the microtubule, leaving the remaining head to walk backwards under its increased load. The weakness of this theory is that it requires a single dynein head to maintain its association with a microtubule that is under tension, while simultaneously shifting between binding sites on the microtubule to allow movement. Even kinesin, single molecules of which have been shown to be functional, cannot do this without using both of its heads¹¹.

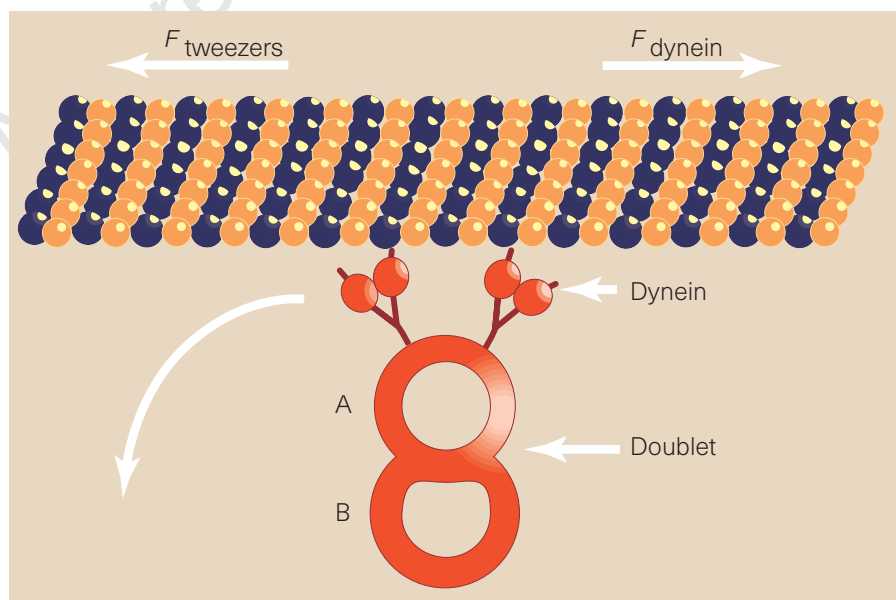


Figure 1 Dynein molecules on a doublet microtubule interact with a microtubule. The doublet is pictured in cross-section and the microtubule is viewed perpendicular to its length. The directions of the forces placed on the microtubule by dynein, and by the optical tweezers used by Shingyoji *et al.*⁴, are shown. The curved arrow indicates how the doublet might rotate due to torque about its attachment to the surface of the slide.

Are there alternatives to these purely mechanical explanations? Although the biochemical milieu of the cell has been removed, these oscillations may nonetheless depend on biochemical modifications. A dynein head has a total of four ATP binding sites¹², but only one of these shows an appreciable rate of hydrolysis. So perhaps strain-dependent changes in the ATP affinity or rates of hydrolysis at the other sites could modulate force generation by dynein.

These new results are an exciting twist in the hunt for the origin of the flagellar beat. Even more exciting is the possibility that the biophysical techniques used will be yet more illuminating when combined with established methods in biochemistry and genetics. □

Alan J. Hunt is in the Department of Molecular, Cellular and Developmental Biology, CB 347, University of Colorado, Boulder, Colorado 80309, USA.

e-mail: hunta@colorado.edu

1. Brokaw, C. *Math. Biosci.* **90**, 247–263 (1987).
2. Murase, M., Hines, M. & Blum, J. *J. Theor. Biol.* **139**, 413–430 (1989).
3. Lindemann, C. *Cell Motil. Cytoskel.* **29**, 141–154 (1994).
4. Shingyoji, C., Higuchi, H., Yoshimura, M., Katayama, E. & Yanagida, T. *Nature* **393**, 711–714 (1998).
5. Kamimura, S. & Kamiya, R. *Nature* **340**, 476–478 (1989).
6. Svoboda, K., Schmidt, C., Schnapp, B. & Block, S. *Nature* **365**, 721–727 (1993).
7. Meyhofer, E. & Howard, J. *Proc. Natl Acad. Sci. USA* **92**, 574–578 (1995).
8. Dutcher, S. *Trends Genet.* **11**, 398–404 (1995).
9. Goodenough, U. & Heuser, J. *J. Mol. Biol.* **180**, 1083–1118 (1984).
10. Gee, M., Heuser, J. & Vallee, R. *Nature* **390**, 636–639 (1997).
11. Hancock, W. & Howard, J. *J. Cell Biol.* **140**, 1395–1405 (1998).
12. Mocz, G. & Gibbons, I. *Biochemistry* **35**, 9204–9211 (1996).

Subtle minds and mid-ocean ridges

Joe Cann

Two-thirds of the surface of the Earth is covered by ocean crust. Each year 3 km² of new ocean floor is created at the spreading plate boundaries, the mid-ocean ridges, requiring that about 20 km³ of molten magma rises each year through the Earth's upper mantle to build new ocean crust. The basic process has been clear for a number of years. But there are some subtle minds around, able to uncover complications in the simplest story. Hence the MELT experiment of the US RIDGE project, one of the largest single Earth science projects ever mounted, results from which are now emerging in print^{1–8} and at conferences.

Understanding of the basic process runs like this. At depth the mantle is hot and soft. This is the asthenosphere, which lies below the tectonic plates. As the plates move apart, the asthenosphere rises to fill the gap between them, and starts to produce a small fraction of silicate melt (magma) as the pressure drops. This magma rises through the mantle to feed magma chambers beneath the axes of the mid-ocean ridges, and from these the ocean crust is constructed.

That much is straightforward. Those subtle minds have created a plethora of models, however, in which the deep mantle may or may not play a part in the process; in which magma may leave the mantle as soon as it has formed, or be retained there to contribute to the dynamics of flow; in which different mechanisms contribute to the focusing of melt from a broad zone of melting in the mantle into the narrow axial magma chamber; or in which melt rises from depth in a narrow zone below the axis. Most of these models make predictions about the current distribution of melt down to about 200 km below the axis of the spreading plate boundary.

These are the issues that MELT was designed to address. The team selected the fastest-spreading mid-ocean ridge system, the East Pacific Rise at 15–18° S (Fig. 1), where the plates are separating at 145 mm yr⁻¹, and where there is a clear axial magma chamber 1–2 km below the sea floor at the spreading axis. The team included seismologists, who use earthquakes and explosions to discover the elastic properties of rocks, and electromagnetic specialists, who use electrical and magnetic fields to estimate the electrical conductivity of the deep Earth. The two types of information are affected by the distribution of magma, which has both zero

rigidity and high electrical conductivity compared with solid rock.

Two lines of seismometers, each 800 km long, were set on the sea floor across the axis of the East Pacific Rise in November 1995 and were recovered six months later, and an array of electrometers and magnetometers was emplaced then and recovered a year later. Within the past month we have had the publication in *Science*^{1–8} of the first results from the seismic experiments; and also the presentation, by R. Evans and A. Chave (Woods Hole Oceanographic Inst.) at a meeting of the American Geophysical Union in Boston, of the initial findings from the electromagnetic experiment.

The MELT seismologists seem most surprised by the asymmetry of their results — there is markedly more melt present beneath the western side of the spreading centre than beneath the eastern side (Fig. 2, overleaf). But this is not so odd, for the area has long been known to be asymmetric. On the west side is the Pacific Superswell, dotted with abundant volcanic cones on top of shallow sea floor, both features implying that the mantle is hotter, while on the eastern side seafloor depths are normal and, hence, so are mantle temperatures. In looking at the results, then, those from the eastern side apply most generally to mid-ocean ridges, whereas those from the western side reflect more the interaction of hot mantle with a spreading centre.

The conclusion from both the seismic and the electromagnetic experiments is that there is a large, tent-shaped zone in the mantle, reaching out to about 150 km away from the spreading axis on the east, and to 350 km away on the west, within which there is a small fraction of melt more or less evenly



Figure 1 The plate tectonic environment of the East Pacific, showing the fast-spreading East Pacific Rise, between the Pacific and Nazca plates. The MELT experiment was performed along the two transects across the East Pacific Rise shown near 20° S, where the total plate separation rate is 145 mm yr⁻¹, the fastest anywhere in the world at the moment. Arrows show relative plate motions.

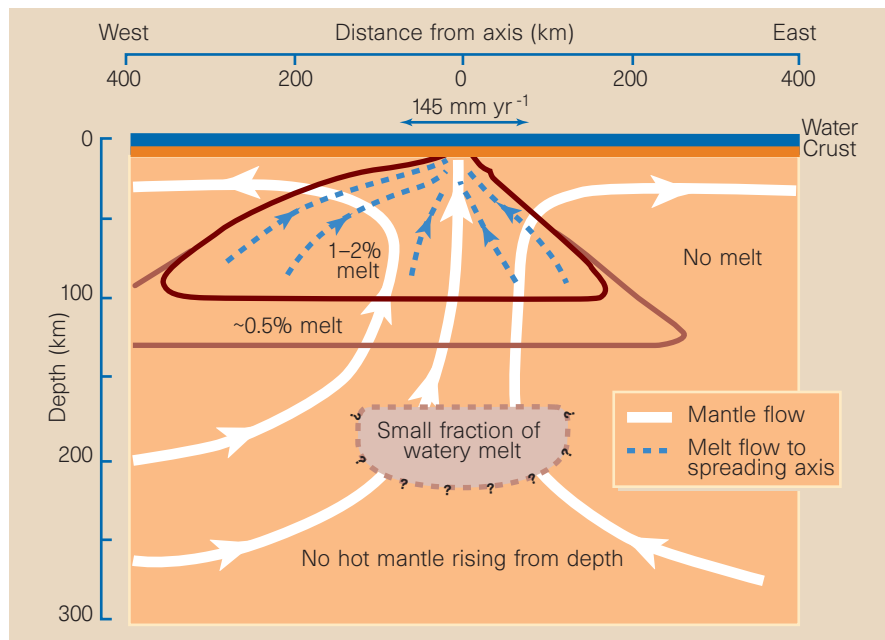


Figure 2 Cross-section through the East Pacific Rise at 17°S showing the initial results of both the seismic and electromagnetic parts of the MELT experiment. The crust is a uniform 6 km thick across the section. Note the asymmetry across the spreading centre, apparently caused by hotter mantle on the west side than on the east, and the asymmetry in mantle flow lines that results. The flow lines of the magma up to the spreading centre simply indicate the overall flow of magma without specifying a precise path. There is no indication from the experiment of deep mantle flow rising beneath the spreading axis. (Modified from Fig. 3 of ref. 1.)

distributed (Fig. 2). Exactly how much there is there depends on how the results are interpreted, which in turn depends on the shape of the droplets of melt and how connected they are to one another. Seismic results indicate that the droplets are elongated, but not stretched into thin films as some theories would require, and that there is about 1–2% of melt in the region. Electromagnetic results show that there can be only a small fraction (less than 1%) of connected melt in this region, because electrical resistivity depends strongly on whether the droplets of melt are joined into a continuous network. Both experiments cannot exclude a narrow zone beneath the spreading axis, less than 10 km wide, much richer in melt, but neither experiment requires it.

The base of the region with 1–2% of melt is defined seismically at about 100 km, but smaller amounts of melt are indicated down to about 130 km. This is important confirmation of deep melting indicated by geochemical reasoning, which had previously been difficult to image seismically. Below this depth, and not detected seismically, is a highly electrically conductive region below 170 km under the spreading axis, which may indicate a very small fraction of well-connected watery melt according to the electromagnetic experimenters (Fig. 2). At greater depths than this there is good seismic evidence that the mantle is no hotter than normal, because the transition zone between the upper and lower mantle is of normal thickness, and not thinner, as it would be if

hot deep mantle were upwelling.

The upper part of the section is intriguing too. Despite the hot mantle close to the axis on the west side, and all of the magma that this contains, one of the seismic experiments shows that the crust is no thicker than normal, and is the same thickness on both sides of the spreading axis. Does this mean that a significant proportion of the magma produced in the hotter area is somehow held back in the hotter mantle? That would be a real surprise.

The interpretation of the seismic experiments in terms of temperatures is not straightforward, as similar seismic anomalies also arise from anisotropy caused by the alignment of crystals in the mantle as it flows. The investigators consider that these problems have been taken into account, but further work is clearly needed to sharpen up these and other conclusions.

Overall, the experiment has been an impressive success, and a great testimony to the power of the complementary evidence to be obtained by combining electromagnetic and seismic data. The results of the experiments strongly favour some models over others. Provided that this area is representative of mid-ocean ridges as a whole, a number of models have bitten the dust in a satisfactorily Popperian way. But models, of course, have a nasty and unPopperian way of bouncing back. □

Joe Cann is in the School of Earth Sciences, University of Leeds, Leeds LS2 9JT, UK.
e-mail: joe@earth.leeds.ac.uk



100 YEARS AGO

Want of accommodation in more than one department of the University museum renders it impossible to carry on satisfactory work. The extracts printed below, from the report of the delegates of the museum, tell of a condition of "hope deferred, which maketh the heart sick." ... After twelve years of fruitless effort to obtain extended accommodation for Honour students, and the means of providing for the increasing number of those working for the preliminary examination ... it is probably quite useless to trouble the delegates with any further application for assistance in this direction. It will be difficult for men of science on the Continent and in the United States to believe that so little encouragement is given to scientific work in the University of Oxford.

From *Nature* 16 June 1898.

50 YEARS AGO

This is one of the most strange, even inexplicable, tragedies in the world of *Nature*. Of the millions of Pacific salmon which enter the rivers to spawn, not one returns to the sea. It is strange ... that they should all die, not one escape, and strangest of all that the change should be from fullest life to death, with no intermission of growing weakness between. The act of spawning, both in the female and in the male, soon brings death, grim and inevitable. ... I once watched two spawned-out, huge fish, in the shallow water of a long, broken rapid. They showed at first just over the lip of the pool above, dorsal and caudal fins breaking as the current forced them down. They drifted a little way, weakly and awkwardly and then, as though suddenly afraid of death, one thrashed its way with furious strength for ten or a dozen yards upstream. The other followed instantly, and for a while they held, side by side again, weaving back and forth until the current caught their sides and swept them down. ... It is strange that they should all die, that life should go out with the spawn; for it does go out with the spawn.

From *Nature* 19 June 1948.

Many more abstracts like these can be found in *A Bedside Nature: Genius and Eccentricity in Science, 1869–1953*, a 266-page book edited by Walter Gratzer. Contact David Plant. e-mail: subscriptions@nature.com

1. The MELT Seismic Team *Science* **280**, 1215–1218 (1998).
2. Canales, J. P., Detrick, R. S., Bazin, S., Harding, A. J. & Orcutt, J. A. *Science* **280**, 1218–1221 (1998).
3. Scheirer, D. S., Forsyth, D. W., Cormier, M.-H. & Macdonald, K. C. *Science* **280**, 1221–1224 (1998).
4. Toomey, D. R., Wilcock, W. S. D., Solomon, S. C., Hammond, W. C. & Orcutt, J. A. *Science* **280**, 1224–1227 (1998).
5. Webb, S. C. & Forsyth, D. W. *Science* **280**, 1227–1229 (1998).
6. Wolfe, C. J. & Solomon, S. C. *Science* **280**, 1230–1232 (1998).
7. Shen, Y., Sheehan, A. F., Dueker, K. G., de Groot-Hedlin, C. & Gilbert, H. *Science* **280**, 1232–1235 (1998).
8. Forsyth, D. W., Webb, S. C., Dorman, L. M. & Shen, Y. *Science* **280**, 1235–1238 (1998).

Holography

Solid information

Hans Coufal

Computers and the Internet have led to an explosion of easily available information. Somewhere, somehow, all of this information must be stored and processed, quickly and economically. Indeed, the decreasing cost of storing data has been a crucial impetus for the information revolution. But magnetic and conventional optical data storage, in which individual bits are stored on the surface of a recording medium, are approaching physical limits beyond which individual bits may be too small to easily write, store or read. Storing information throughout the volume of a medium, rather than just on its surface, is an intriguing high-capacity alternative. One such volumetric approach, holographic data storage, was conceived decades ago. But it tends to be fragile — reading out information tends to destroy what is written. On page 665 of this issue¹, Buse *et al.* describe a way to make holographic storage much more durable.

In holographic data storage², an entire page of information — a million or so data bits — is stored at once as an interference pattern within a thick, photosensitive optical material. This is done by crossing a so-called object wave (typically a laser beam, on which the information has been imprinted by a suitable light modulator, such as a small liquid-crystal TV screen) with a reference beam carrying no added information. A large number of these interference gratings or patterns can be stored in the same piece of material, as long as they are distinguishable by their direction or spacing. That can be accomplished by changing the angle between the object and reference wave or by changing the laser wavelength. A page is read out by illuminating the material with the reference wave that was used to store that page: the wave is scattered in such a fashion that the object beam is reconstructed. The theoretical limit for the storage density of this technique is around tens of terabits per cubic centimetre, although the densities that have been demonstrated so far are three orders of magnitude smaller, comparable to a standard disk drive.

In addition to high storage density, holography promises fast access times, because the laser beams can be moved rapidly without inertia, unlike the actuators in

disk drives; and also high data rates, because of the inherent parallelism of its page-wise storage and retrieval. But despite all of these obvious advantages — and the availability of cheap components such as liquid-crystal displays for spatial light modulators and CCD camera chips from video recorders for detector arrays — no commercial holographic storage products are available today. Why? The chief hurdle is the lack of a suitable recording medium. Despite the efforts of academic and industrial laboratories around the globe, there remains a critical problem. Recording or reading out a page tends to partially erase any holograms recorded earlier.

Non-destructive read-out requires some process that renders the stored information permanent, similar to fixing a photograph during film development. Depending on the specific mechanism used to record the holograms, fixing could involve applying an electrical field or annealing with heat. But the required high voltages, ovens or high-power lasers would be expensive, and possibly troublesome in a consumer device. Such fixing techniques would also make it much more difficult to verify information immediately after it had been written, as is typically done with magnetic and optical recording today.

One might avoid fixing by using a material whose optical properties are wavelength

or optical-power dependent. The simplest method is to record a hologram at a wavelength for which the material is photosensitive (green, for example) and reads it out at another wavelength at which the material is not sensitive to light (infrared, for example). This is like a photographer using a red light in the darkroom to observe pictures being printed using white light. Using a different read-out wavelength doesn't work well, however, because it changes where the individual pixels appear on the detector array, introducing errors that are difficult to avoid. Another approach uses a high-power laser — data are written at high power, taking advantage of nonlinear effects, and read back with low-power light pulses well below the writing threshold³. Unfortunately, high-power, pulsed lasers are complex and will be expensive for the foreseeable future.

Still another approach is a one-two punch using two different wavelengths^{1,4–6}. In the device described by Buse *et al.*¹, an ultraviolet light source excites electrons in the recording medium from a low-energy state (on a manganese ion 'trap') to an intermediate energy state (on an iron ion) via the conduction band — a process called gating or biasing the material. Red light then has enough energy to write the hologram in the material by exciting the electrons from the intermediate state up into the conduction band; these electrons quickly diffuse the relatively short distance into the deeper manganese traps (see Fig. 1 on page 665). The electric fields between these trapped electrons and the ions left behind establish localized changes in the refractive index of the host material, forming the hologram's interference grating.

Such a hologram is stable because further illumination at the writing wavelength (red) cannot excite the trapped electrons back into the conduction band. For read-out, light at the writing wavelength is scattered from

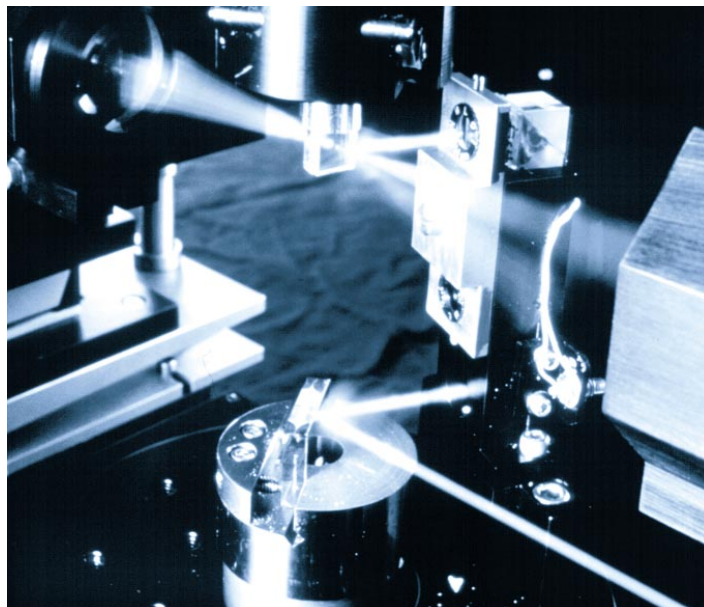


Figure 1 Light writing: object and reference laser beams intersect in a doped LiNbO₃ crystal to record a hologram containing 1 Mb of information. (Courtesy of IBM Almaden Research Center.)

these localized inhomogeneities, and the stored wavefront is reconstructed. As the first light source need not be coherent, or even very narrowly peaked in wavelength, light-emitting diodes or lamps with colour filters can be used, instead of an expensive laser emitting visible light. And as a by-product, the holographic medium can be bulk erased by an exposure to the gating light.

As mentioned at the beginning, holograms are easily erased by those recorded later in the same volume. This can be alleviated partially by anticipating the partial erasure: the first holograms can be recorded for much longer times, thus achieving in the end a relatively equal intensity for all recorded holograms. Another approach would be to refresh the information periodically, as is done in dynamic 'DRAM' semiconductor memory. For both solutions, the crucial property is the '*M* number' — the ratio of the speed at which holograms are recorded and erased. In the new work¹, *M* would reach about 7 for a 1-cm-thick sample. That means that erasure is seven times slower than the recording speed — a very desirable ratio.

A practical material must meet many other criteria, too. High sensitivity at the writing wavelength would eliminate the need for a large, expensive, high-power laser.

In the present system, the high *M* was accomplished without sacrificing the sensitivity to the point that the results would be of only academic interest.

This study is a step towards a practical holographic storage device. But it is by no means the last one. A truly useful holographic recording medium must also have the right optical and mechanical properties, and be insensitive to temperature and cheap to make. Many more advances are necessary before we can buy a device at a price that is competitive with established data storage technologies. Moreover, holography fans should not overlook the fact that its well-entrenched competitor technologies are improving rapidly — by about 60% a year. □

Hans Coufal is at the IBM Almaden Research Center, 650 Harry Road, San Jose, California 95120-6099, USA.

e-mail: coufal@almaden.ibm.com

1. Buse, K., Adibi, A. & Psaltis, D. *Nature* **393**, 665–668 (1998).
2. Psaltis, D. & Mok, F. *Sci. Am.* **273**, 70–76 (1995).
3. von der Linde, D., Glas, A. M. & Rodgers, K. F. *Appl. Phys. Lett.* **25**, 155 (1974).
4. Bai, Y. S. & Kachru, R. *Phys. Rev. Lett.* **78**, 2944–2947 (1997).
5. Guenther, H., Wittmann, G., Macfarlane, R. M. & Neurgaonkar, R. R. *Opt. Lett.* **22**, 1305–1307 (1997).
6. Lande, D., Orlov, S. S., Akella, A., Hesselink, L. & Neurgaonkar, R. R. *Opt. Lett.* **22**, 1722–1724 (1997).

Palaeobiology

101 uses for fossilized faeces

Peter Andrews and Yolanda Fernandez-Jalvo

Fossils are familiar objects by which we can trace the evolution of life on Earth and document past climatic events. Less familiar are the so-called trace fossils — traces that once-living animals have left behind them as a result of something they did during life. On page 680 of this issue, Chin *et al.*¹ describe a piece of fossilized dung from Canada that, from its large size and context, can be reliably assigned to a very large carnivorous species of dinosaur, possibly *Tyrannosaurus rex*.

Trace fossils are examples of fossilized behaviour, and, if they can be assigned to a fossil species, this tells us something about the behaviour of that species. For terrestrial vertebrates, the most common trace fossils are footprints, and many examples of dinosaur², mammal³ and even human⁴ footprints have been found. Not as well known — and for obvious reasons much less glamorous — are fossilized faeces or pellets, usually referred to as coprolites. There are many examples of coprolites from mammalian or avian carnivores⁵, but few have ever been found that could be assigned with any certainty to carnivorous dinosaurs.

The contents of coprolites are the main issue here. These provide direct evidence

about how and where the coprolite-producer fed. Ideas on dinosaur behaviour have changed dramatically, and the new discovery not only changes our thinking on how carnivorous dinosaurs ate their food, but it gives us clues about the association between



Figure 1 Photomicrograph of a rodent femur digested by a crocodile. Predictions can be made about bones found in fossilized faeces (coprolites), such as those discovered by Chin *et al.*¹, by comparison with bones in the faeces of modern-day crocodiles. Crocodiles have low enzymatic activity during digestion, and their gastric juices are almost pure hydrochloric acid. This results in very aggressive destruction of the inorganic fabric of bone tissues, so only organic collagen fibres remain.

predator and prey. Predators are very effective in sampling their environment, but they are selective, only taking prey of particular size or type. They may transport prey from one place to another, so the more mobile the predator the further a bone may have been removed from where the prey originally lived. The same goes for pollen grains that may be inside the coprolite. These could be derived from the diet of the herbivorous prey, or they may have been trapped by the sticky surface of the coprolite when it was still fresh⁶. The species content of coprolites therefore depends on the hunting preferences of the predator, assuming that it is the product of hunting as opposed to scavenging. On the other hand, many cases are 'now' where animals turn up in predator scats in places where their presence had never even been suspected.

The theropod coprolite described by Chin *et al.* is extremely large, and it contains the remains of a sub-adult ornithischian dinosaur. Finding the remains of food in faeces or pellets, or sometimes in the stomach of a well-preserved specimen⁷, gives a direct insight into what a predator had been eating. Although this is interesting in itself, it becomes still more fascinating when the bones of the prey are examined. All predators modify their prey, either by the way in which they ingest it or by the way that they digest it. By analogy with living reptiles, dinosaurs would have been expected to ingest their prey in relatively large pieces⁸, so breakage of bone should not be extensive. But this was apparently not the case. The authors found that the theropod coprolite contained many bone fragments, ranging from a few millimetres to less than 100 micrometres in length. This bone was evidently crushed — one could say pulverized — during ingestion, which implies a considerable degree of food preparation before it was swallowed. This may have been similar to the repeated biting of small prey animals by crocodiles, the largest living reptiles, before they swallow them.

But just the reverse happens after ingestion. By analogy with crocodiles⁹, we can predict that ingested bone would have been extremely heavily digested in carnivorous theropods (Fig. 1). Although some of the small (less than 1 mm in diameter) bone fragments from the coprolite seem to be completely unstructured (Fig. 3 on page 681), the photomicrograph of one of the larger fragments (Fig. 2 on page 680) shows the internal histology to be intact and unmodified. In fact, this bone is in such apparently pristine condition that it might be worth looking for DNA in it. The edges of all the bone fragments are rounded to some degree, but this is little more than would be expected in most mammalian carnivores⁵, and nothing like the degree of modification seen in crocodile scats. ►

The contrast between the extent of breakage and the low level of digestion is the most surprising finding from the newly discovered coprolite. Not only does it throw new light on the ability of large carnivorous dinosaurs to break down the bones of their prey in their mouths, but it also tells us something about the physiology of digestion once the prey animal has been swallowed. All of that from a piece of dung! □

Peter Andrews and Yolanda Fernandez-Jalvo are in the Department of Palaeontology, Natural

History Museum, London SW7 5BD, UK.

e-mail: P.Andrews@nhm.ac.uk

1. Chin, K., Tokaryk, T. T., Erickson, G. M. & Calk, L. C. *Nature* **393**, 680–682 (1998).
2. Gillette, D. D. & Lockley, M. G. *Dinosaur Tracks and Traces* (Cambridge Univ. Press, 1989).
3. Kordos, L. *Geol. Hungarica Palaeontol.* **46**, 257–415 (1985).
4. Leakey, M. D. & Hay, R. L. *Nature* **278**, 317–323 (1979).
5. Andrews, P. *Owls, Caves and Fossils* (Chicago Univ. Press, 1990).
6. Scott, L. *Quat. Res.* **28**, 144–156 (1987).
7. Charig, A. J. & Milner, A. C. *Bull. Nat. Hist. Mus. Lond. (Geol.)* **53**, 11–70 (1997).
8. Johnson, C. R. *Zool. J. Linn. Soc.* **52**, 315–336 (1973).
9. Fisher, D. C. *Paleobiology* **7**, 262–275 (1981).

HIV

Envelope's letters boxed into shape

John P. Moore and James Binley

The groups of Sodroski and Hendrickson have put their distinctive stamps on the HIV-1 envelope by delivering a package of information (on pages 648¹ and 705² of this issue, and in this week's *Science*³) on the crystal structure of the gp120 surface glycoprotein. These findings complement reports^{4,5} on the structure of gp41, and have important implications for virology, immunology and vaccine development.

The two groups crystallized the 'core' of the gp120 molecule from the HxBc2 laboratory strain as a ternary complex with its primary receptor (CD4 domains 1 and 2) and an anti-gp120 antibody fragment called 17b, which partially mimics the HIV-1 co-receptor (CCR5/CXCR4). Although seemingly drastic, the modifications necessary to crystallize the gp120 core preserve its antigenic integrity⁶, and they do not seriously affect the value of the structural information obtained.

Much of what has been surmised about the topology of gp120 by biochemical⁷, mutagenic^{8,9} and immunochemical¹⁰ techniques is now confirmed by the crystal structure. Additional surprises are, however, nicely discussed by the authors^{1,2}. These include the existence of a 'silent face' — a large, previously unsuspected surface — and the nature of the two-domain gp120 structure, which provides a natural mechanism for receptor-induced conformational changes. Residues from several regions of the gp120 core are brought together to form the broad area that associates with CD4 but, unexpectedly, many of the CD4 contacts are made using the peptide backbone of gp120 amino acids, not their side chains. Because antibody epitopes usually involve side chains, this device allows HIV-1 to alter the residues that form the CD4 binding site, without penalty to receptor binding, while changing the antigenic structure of the site to evade receptor-blocking antibodies. Nonetheless, there are also critical contacts with a few conserved side chains, including a

'knob-and-socket' interaction involving the unusually protuberant Phe 43 of CD4 and a receptive hole in gp120. This, and other cavities revealed in the surface of gp120, provides a target for inhibitors of receptor binding.

The region of gp120 that binds the CXCR4 co-receptor is also revealed on the crystal structure, in surrogate, by the residues that contact (or are close to) the 17b antibody fragment. These residues are located within the highly conserved stem of the V1/V2 structure, near the base of the V3

loop, and in other regions folded into proximity. The similarities between the binding sites for the CXCR4 and CCR5 co-receptors are likely to outweigh their differences, as shown by Rizzuto *et al.*³. They used the HxBc2 gp120 crystal structure to guide the design of gp120 mutants from YU2, a primary HIV-1 that uses CCR5. The amino-acid substitutions that knock out CCR5 binding in the YU2 gp120 also destroy the 17b epitope, and this makes sense when the position of these residues is modelled onto the structure of the HxBc2 gp120.

The precise nature of the CXCR4-binding site can be only partially inferred from the crystal structure at this stage. But many of the residues involved in CCR5 binding are likely to be important for contacting CXCR4. Moreover, we can assume that there will be additional contacts between CXCR4 and the V3 loop (and, perhaps, some V2 residues). These could involve the positively charged residues characteristic of the V3 loop of viruses that use CXCR4, and negatively charged residues on CXCR4. Perhaps the best way to visualize the co-receptor interactions of gp120 is that a high-affinity site for CXCR4 is created on the background of a conserved CCR5/CXCR4 site by sequence changes in the variable loops (Fig. 1). These changes may also affect the geometry of the CCR5

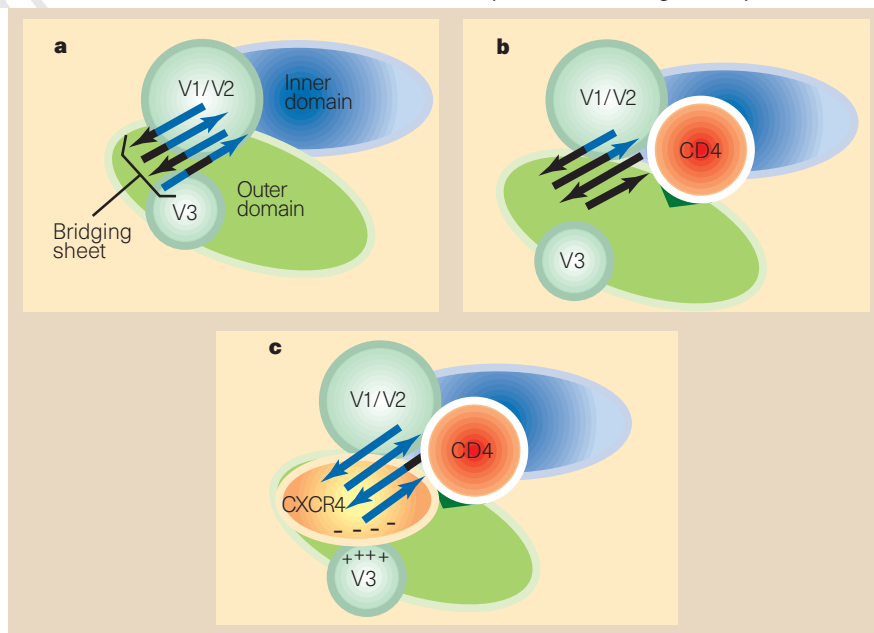


Figure 1 Proposed conformational changes during binding of gp120 to its receptor (CD4) and co-receptor (CXCR4/CCR5). These changes are based on the crystal structure of gp120 reported by Kwong *et al.*¹, Wyatt *et al.*² and Rizzuto *et al.*³. a, The bridging sheet (arrows), which spans the outer and inner gp120 domains, is made of residues in the stem of the V1/V2 loop and the CD4 region, and has not been shown to exist before binding to CD4. b, CD4 binds to a region that forms a crevice in gp120, triggering a conformational change that moves the V1/V2 loops away from the bridging sheet, and which may also move the inner and outer domains relative to one another, exposing the bridging sheet. c, The flexibility of CD4 allows the viral envelope to approach the target cell membrane, where the CXCR4 co-receptor interacts with the bridging sheet. Negatively charged regions of CXCR4 complement positive charges near the base of the V3 loop (this is speculative). The charge or conformation of the V3 loop may determine whether the CXCR4 or CCR5 co-receptor is used.

site, occluding it in gp120s (such as HxBc2) that bind only CXCR4.

Importantly, creation or exposure of the highly conserved co-receptor-binding site requires that gp120 first binds CD4 (refs 11–13). This is another way for HIV-1 to evade humoral immunity — by the time the co-receptor site is ready to bind CCR5 or CXCR4, the virus is already attached to CD4. Steric constraints will hinder access of antibodies to the co-receptor site under these conditions, explaining why primary isolates are poorly neutralized by the 17b antibody^{2,3}. The CD4-induced conformational changes in gp120 involve movement of the V1/V2 structure and, to a lesser extent, the V3 loop, away from the underlying co-receptor-binding site¹¹. Although these variable loops are not present on the crystal structure, they have been modelled¹⁰ as a protuberance above the gp120 core. One way to view them is as an umbrella that shields the critical regions of gp120 from the rain of antibodies thrown at it by the humoral immune response; if a neutralizing antibody succeeds in binding to the variable loops, the virus will simply mutate the non-essential residues involved, and escape.

The virus has additional protection from humoral immunity by the extensive glycosylation of gp120. The authors^{1–3} modelled many of the glycans onto the crystal structure, clearly revealing how they shield receptor-binding regions of the peptide backbone from antibodies. This makes sense from the virus's perspective — with rare exceptions, HIV-1 is neutralized by inhibition of its attachment to cellular receptors¹⁴. The same protective devices will also reduce the binding of gp120 to the immunoglobulin-like B-cell receptor, meaning that HIV-1 can also limit the production of neutralizing antibodies in the first place. Throw in observations that some strains of HIV-1 can even use anti-gp120 antibodies to increase their ability to fuse with host cells¹⁵ — presumably by occupying one of the three subunits of an assembled envelope glycoprotein trimer and inducing structural changes in the other two — and the war between HIV-1 and the humoral immune system takes an even more perverse twist.

The trimeric nature of the assembled gp120–gp41 complex can only be inferred from the crystal structure because the inter-subunit contacts are between the gp41 moieties. But there is really only one way for all the components to fit together^{1,2}. The immunogenicity and antibody reactivity of the assembled complex are even less than those of the gp120 monomer, perhaps because of steric considerations^{16,17}, and this provides yet another level of protection — the immune system is decoyed into making antibodies to disassembled gp120 that are

poorly reactive, and hence ineffective, with virions. These protective measures may reduce HIV-1 infectivity *in vivo*, but they provide an overall advantage in the face of the immune response. *In vitro*, HIV-1 can afford to discard some of its protective armour, increasing its ability to bind receptors and infect its target cells at the (now irrelevant) expense of becoming neutralization sensitive¹⁸.

So what can be done to overcome the defences of HIV-1, given that an antibody response may be necessary to supplement vaccine-induced cellular immunity? There seems little to be gained by continuing to use simple gp120 subunits of whatever strain, alone or in combination. Antibodies elicited by such proteins play into the virus's hands because they attack its defences head-on. If an arrow bounces off a tank, why use a quiver-full of the same design? Instead, we need to use the crystal structure to design a smart bomb with armour-piercing capacity, perhaps by modifying the antigenic structure of gp120. Already, there are indications that this may be possible. When glycosylation sites were deleted¹⁹ from the V1/V2 loops of the simian immunodeficiency virus gp120, not only was a neutralization-sensitive virus created, but the immunogenicity of the mutant virus was altered so that a better immune response was raised to the wild-type virus. Similarly, removing the V1/V2 loops from HIV-1 gp120 renders the conserved regions underneath more vulnerable to antibodies^{11,20}, although it is not yet known whether this will translate into improved immunogenicity. These and other approaches that will be stimulated by the new information on the structure of gp120 are part of the way ahead on the long road to developing an HIV-1 vaccine. □

John P. Moore and James Binley are at the Aaron Diamond AIDS Research Center, The Rockefeller University, New York, New York 10021, USA.

e-mail: jmoore@adarc.org

1. Kwong, P. D. *et al.* *Nature* **393**, 648–659 (1998).
2. Wyatt, R. *et al.* *Nature* **393**, 705–711 (1998).
3. Rizzuto, C. *et al.* *Science* **280**, 1949–1953 (1998).
4. Weissenhorn, W., Dessen, A., Harrison, S. C., Skehel, J. J. & Wiley, D. C. *Nature* **387**, 426–430 (1997).
5. Chan, D. C., Fass, D., Berger, J. M. & Kim, P. S. *Cell* **89**, 263–273 (1997).
6. Binley, J. M. *et al.* *AIDS Res. Hum. Retroviruses* **14**, 191–198 (1998).
7. Leonard, C. K. *et al.* *J. Biol. Chem.* **265**, 10373–10382 (1990).
8. Olshchinsky, U. *et al.* *J. Virol.* **64**, 5701–5707 (1990).
9. Helseth, E. *et al.* *J. Virol.* **65**, 2119–2123 (1991).
10. Moore, J. P. & Sodroski, J. *J. Virol.* **70**, 1863–1872 (1996).
11. Wyatt, R. *et al.* *J. Virol.* **69**, 5723–5733 (1995).
12. Wu, L. *et al.* *Nature* **384**, 179–183 (1996).
13. Trkola, A. *et al.* *Nature* **384**, 184–187 (1996).
14. Ugolini, S. *et al.* *J. Exp. Med.* **186**, 1287–1298 (1997).
15. Sullivan, N. *et al.* *J. Virol.* (in the press).
16. Moore, J. P. *et al.* *J. Virol.* **69**, 101–109 (1995).
17. Burton, D. R. & Montefiori, D. *AIDS* **11**, S587–S598 (1997).
18. Moore, J. P. & Ho, D. D. *AIDS* **9**, S117–S136 (1995).
19. Reitter, J. N., Means, R. E. & Desrosiers, R. C. *Nature Med.* **4**, 679–684 (1998).
20. Cao, J. *et al.* *J. Virol.* **71**, 9808–9812 (1997).

Daedalus

Thermal noise

How to get rid of our mounting piles of organic waste? Oxidation is the obvious reaction for the job; but burning in air generates highly unpopular smoke. Water-based oxidation would be far better. Sadly, it either needs ferocious reagents, such as fuming nitric acid, or extremely high temperatures and pressures, as in supercritical aqueous oxidation. Daedalus is looking for another way.

He notes that sonolysis, subjecting a reaction to intense high-frequency sound, can speed it up hundreds of times. The violent pressure-swings of the sound cause the liquid to cavitate, that is, to form tiny transient bubbles of vacuum. Their collapse produces vast temperatures and pressures; these create energetic free radicals which speed the reaction.

Sonolysis can certainly accelerate the oxidation of organics in solution. But Daedalus wants to destroy solids as well — old newspapers, plastic rubbish, food residues, discarded clothing, and the rest of our organic detritus. He points out that bubbles form easily on solid surfaces, especially irregular ones. Hence the 'boiling stones' used by chemists to aid smooth boiling, and the cavitation suffered by ships' propellers. A propeller can stir the water violently enough to cause cavitation; the bubbles form right on the metal where they can do the most damage.

In principle, therefore, a suspension of solid waste should oxidize if stirred with sufficient vigour — provided the waste itself was used as the stirrer. Now an object suspended in a conducting liquid threaded by a magnetic field experiences a force when a current is passed through the liquid: a sort of differential motor effect. So Daedalus will oxygenate his rubbish suspension, put it in a strong magnetic field, and pass high-frequency a.c. through it. The violent vibration of the solids against the surrounding water will cause cavitation at their interface. Bubbles will form and collapse on the solid surfaces, exactly where they are needed; the suspended waste will erode and oxidize rapidly.

Daedalus's waste-cavitation plant will suspend its shredded waste in air-saturated sea water, the cheapest oxidizing conducting solvent. The rubbish will simply fizz away to gas and ash, and the sterile effluent will be returned to the sea. The process should work on domestic sewage, too. Those lazy seaside towns that just pump the stuff out to sea will not even have to change their outfall pipe.

David Jones

Obituary

Pere Alberch (1954–98)

Synthesizer of development and evolution

A bold and creative scientist has left us before he could fulfil his vast potential. Pere Alberch, who was inspired to attempt what had been an elusive goal for many, a synthesis of development and evolution, died in his sleep on 13 March from heart failure. He was only 43.

Dreams of that synthesis, which would explain how biological form evolves, have a long history. Pre-Darwinian French biologists, notably Geoffroy Saint-Hilaire, laid the early groundwork; Charles Darwin saw the potential; and Ernst Haeckel took on the task. But it was too soon for all of them. Other prominent scientists pursued the goal — William Bateson and Thomas Hunt Morgan in the early days of genetics, and later Gavin de Beer, C. H. Waddington and I. I. Schmalhausen.

The most recent era opened in 1977, with the appearance of S. J. Gould's *Ontogeny and Phylogeny*, published by Harvard University Press, which presented Gould's views on how development (ontogeny) and evolution (phylogeny) might be related. Reading this treatise, Alberch, George Oster and I saw how Gould's formulation could be extended in new directions if his model were to be clothed in mathematical form, so we invited him to join us in writing a paper. Published in *Paleobiology* (5, 296–317; 1979), the paper stimulated much research, became a citation classic, and provided a point of departure for Alberch's career.

Alberch was fascinated by morphogenesis and the evolution of body plans. Some criticized him for not being 'more genetic', but what interested him was the genesis of structure and form. He turned his attention to the mechanics of morphogenesis, for he felt that this was the black box between gene and phenotype that must be unlocked. He saw that an eclectic approach was required that combined concepts from engineering, mathematics, genetics and evolution. He mastered them all, and brought them together to forge a unique perspective on development and evolution. He blended comparative (phylogenetic) and experimental methodologies, and designed clever and direct experiments, focused sharply on specific questions such as factors controlling the numbers of digits in frogs versus salamanders.

Born into an old Catalan family in 1954, Alberch had a passion for natural history which was fostered by his mother



and grandfather. Early on, he affiliated himself with the Museu de Zoologia and the zoo in Barcelona, and began research projects that led to his first two publications in 1973. He left Spain to pursue his interests in herpetology, first at the University of Kansas, and then at Berkeley, where he joined my group in 1976. He soon realized that he would need more grounding in mathematics and theory, and so he sought out and began working with George Oster. George and I served as co-supervisors for Alberch's doctoral dissertation. He worked fast, and by the time he was 25 he was assistant professor at Harvard University and assistant curator in the renowned Museum of Comparative Zoology.

At Harvard, Alberch quickly established a productive research programme and gained funding for his own research as well as for development of the magnificent collection of amphibians and reptiles under his care. He attracted promising students, and several have become prominent in the field now known as 'devo-evo' — Neil Shubin, Cliff Tabin, Annie Burke and Chris Rose among them. Always ready to challenge authority, Alberch followed his vision, both personal and professional, without regard to the consequences. He could be charming, but also outrageous, and he drew attention by both his personality and his innovative research.

By 1985 he had revised some of his own positions from earlier years, just as others were beginning to perceive their potential. In a series of research papers, and in a remarkable essay, "The logic of monsters" (*Geobios* mém. spéc. 12, 21–57; 1989), he showed more fully how development and evolution could be studied. He had faith in his view of evolution, and although he

could be sharply critical, he always maintained an open mind. Perhaps these are not traits that non-tenured faculty at Harvard should emulate, for Harvard did not know what to make of him. What was he? Herpetologist? Developmental biologist? Evolutionary biologist? In the end Harvard failed to recognize what it had, and Alberch departed for Spain to initiate a new phase in his career.

In 1989 he became the director of the Museo Nacional de Ciencias Naturales in Madrid, and a research professor in the Spanish research council (CSIC). There he established his laboratory and threw himself into the tasks of directorship. He reorganized research departments, encouraged modernization of the collections, and was the driving force behind new and stimulating exhibits. He worked closely with colleagues throughout Europe to modernize museum science, all the while carrying on his research on salamander biology as a case study in evolution and development.

In many ways Alberch was a renaissance figure. He was intensely interested in contemporary art, especially that of his native Catalonia (his personal favourite was Joan Miró). He developed a reputation as an art critic and collector, and displayed a remarkable ability to find and help develop new talent in photography and painting.

Pere became seriously ill in 1995 and left his directorship. Recently, he felt well enough to return to research and teaching, and was planning a move to Valencia when he died. For one so young he leaves an imposing intellectual legacy. His contributions were widely recognized by his peers and he was honoured with special appointments at Cambridge University, Uppsala University, at various institutions in Argentina, and at Berkeley. Students were attracted to him because of his dynamic and irreverent style — and his enjoyment of controversy. He believed strongly that science advances by establishing bold initiatives and by testing explicit hypotheses.

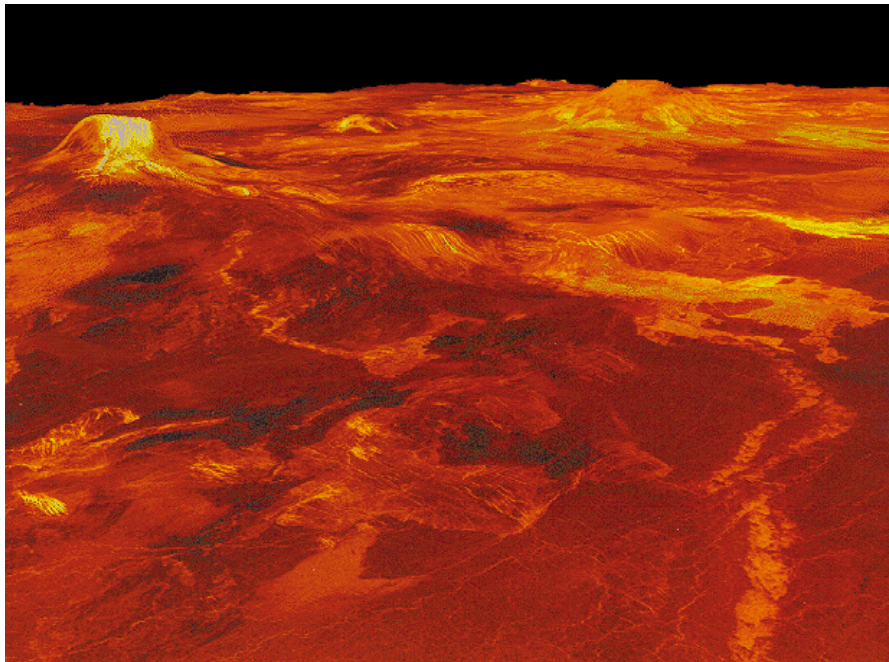
Although Pere Alberch's studies in evolution and development were highly influential and earned him an international reputation, those who knew him well believe that circumstances prevented the full realization of his talents. The best was yet to come, and Pere's early death has cost the scientific community dearly.

David B. Wake

David B. Wake is in the Museum of Vertebrate Zoology, University of California, Berkeley, California 94720-3160, USA.
e-mail: wakelab@uclink4.berkeley.edu

Venus's voyeurs

Pictures of far-off planets sent back to Earth by spacecraft allow us all to be armchair explorers. The rendering of these images in a form that we can see involves choices that would be familiar to any traditional landscape painter.



Perspective view of volcanoes at Western Eistla Regio on Venus: from the Magellan mission in 1991.

Martin Kemp

When images of Jupiter, beamed back by a 'suicide probe' from the spacecraft Galileo, were published in Britain in *The Independent* newspaper on 7 June 1997, the text began: "It may look like a Turner painting..." We may recall that Turner himself in the early nineteenth century was striving to express the awesome infinity of a scale-less cosmos characteristic of Romantic science, art and literature.

Spectacular examples of what appear to be sublime landscapes seen with our normal visual apparatus are provided by the views of Venus generated from the huge quantities of data transmitted from the Magellan spacecraft in 1991–94.

Such is the nature of the atmosphere of Venus that it cannot be penetrated by rays within the normal visible spectrum. The 'visual' technique involved synthetic aperture radar scanning, in which a radar beam was directed diagonally downwards and sideways from Magellan's rapid path. The sideways look, akin to that of side-scan sonar in oceanic surveys, was complemented by radar altimetry, using a beam pointed directly downwards.

Some of the effects of 'seeing' by radar are visually anomalous. For example, highly reflective surfaces return little or no reflections directly back to the laterally positioned transmitter/receiver, and therefore register

as 'black' — in contrast to their reflective lustre when viewed by the eye in normal lighting conditions.

To achieve the beautiful landscapes, those synthesizing the images from the side-scans and altimetry need to take a series of overtly pictorial choices which affect every aspect of the rendering — from the basic decision to map the data using orthodox perspectival scaling, to detailed choices of what 'false colours' should be used and how they are to be modified by distance.

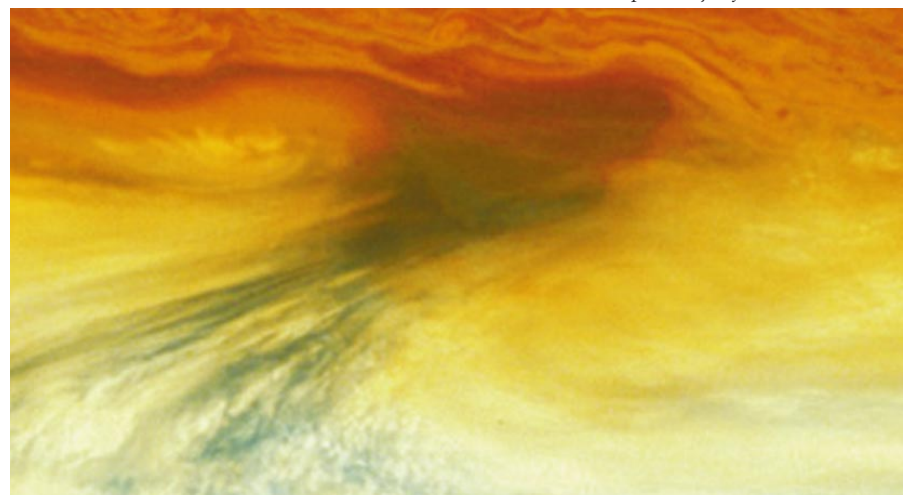


Image of the atmosphere of Jupiter provided by the Galileo spacecraft.

The resulting pictures are masterpieces by largely anonymous masters of extra-terrestrial landscape painting. The picturesque and sublime beauties of our 'sister' planets are revealed according to the same kinds of pictorial rendering that have come to certify the visual delights of our terrestrial habitat.

But why are these images produced, when our scientific understanding of the planets does not appear to be significantly advanced by such seductive extravagance?

I think there are various reasons. One relates to the human impulses of those who undertake incredible journeys of exploration. The naming of the craft 'Magellan', after the great Renaissance voyager, provides a strong clue. The desire of intrepid explorers, ancient and modern, to see the wonders of the newly discovered terrains with their own eyes is irresistible.

Another explanation lies in the satisfying of the voracious public need for pictures of discoveries within a culture that demands cascades of visual novelty.

A related reason is that the enormous budgets for space exploration need to be justified by space agencies and their political masters. Some kind of spectacular public output is required if the enterprises are to continue.

The professional explorers and the public audience share a series of predilections long built into our visual culture. We become by electronic proxy armchair tourists and aesthetic voyeurs in extraordinary voyages of stellar exploration. □

Martin Kemp is in the Department of the History of Art, University of Oxford, 35 Beaumont Street, Oxford OX1 2PG, UK.

e-mail: martin.kemp@trinity.oxford.ac.uk

Mechanisms for memory types differ

The formation of long-term memory takes several hours^{1–3}, during which time memories rely on short-term systems^{1,2,4,5}. For over 100 years¹, the main unanswered question of memory research has been whether short-term memory is a necessary step towards long-term memory^{4,5}, or whether they are separate processes^{1,2}. Here we report four treatments that block short-term memory while leaving long-term memory intact, showing that these memory systems are separate to some degree.

The treatments we used here all alter long-term memory when infused into the CA1 sub-region of the hippocampus or the entorhinal cortex of rats that have been trained to perform certain behaviours^{6–8}. These treatments were the glutamate AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid) receptor antagonist 6-cyano-7-nitroquinoxaline-2,3-dione (CNQX, 0.5 μ g), the GABA_A (γ -aminobutyric acid type A) receptor agonist muscimol (MUS, 0.5 μ g), the serotonin 1A receptor agonist 8-hydroxy-2-(di-*n*-propylamino) tetralin (DPAT, 2.5 μ g) and the serotonin 1A receptor antagonist 1-(2-methoxyphenyl)-4-(4-(2-phthalimido)) butylpiperazine (NAN, 2.5 μ g).

We implanted 30-gauge guides bilaterally 1 mm above the dorsal CA1 region of the hippocampus ($A = 4.3$, $L \pm 4.0$, $V 3.4$) or 1 mm above the surface of the entorhinal cortex ($A = 7.0$, $L \pm 5.0$, $V 8.4$) of Wistar rats (240–300 g) under deep thionembutal anaesthesia; stereotaxic coordinates are given in millimetres according to ref. 9. After recovery, the rats were placed on a platform that was 25 cm high and 7 cm wide. This platform faced a 43 $\times$ 25 cm grid of stainless steel bars, spaced 1.0 cm apart and of 0.1 cm in width. The platform was used for inhibitory avoidance training. We measured how quickly the rats stepped down onto the grid with all four paws. Once on the grid, the rats received a scrambled electric shock to their paws of 0.3 mA for 1 second. They immediately received bilateral infusions of 0.5 μ l saline, a vehicle (20% dimethylsulphoxide) or a drug. MUS and DPAT were dissolved in saline, and CNQX and NAN were dissolved in the vehicle. Infusion cannulae protruded 1.0 mm beyond the guides. Infusion procedures and verification of cannula placement were performed as described^{6–8}.

There were two main experiments. In the first, animals were tested twice to see whether they had retained the memory of the electric shock: once at 1.5 hours after training, to measure short-term memory, and once at 24 hours after training, to measure long-term memory (Fig. 1a, b). In the second experiment, we tested the animals 1.5, 3.0 and 4.5 hours after training (Fig. 1c, d). Test

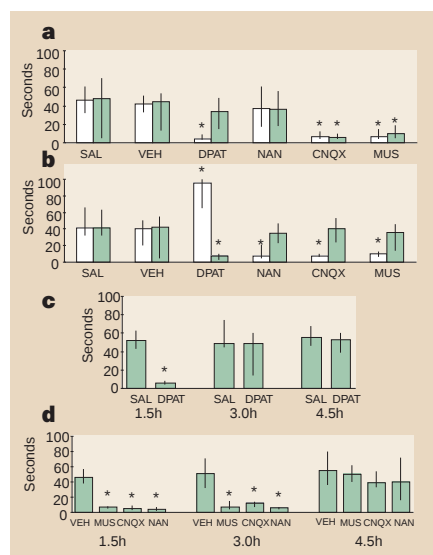


Figure 1 Data are expressed as median (interquartile range) test session latency in the step-down inhibitory avoidance task. **a**, **b**, Test latency was measured at 1.5 h after training (short-term memory; white columns) and at 24 h after training (long-term memory; shaded columns). Each animal was tested twice, first for short-term and then for long-term memory. Rats received immediate post-training bilateral infusions of the named compounds in the dorsal CA1 subregion of the hippocampus (**a**) or in the entorhinal cortex (**b**). SAL, saline; VEH, vehicle. **c**, The effect of intrahippocampal DPAT on short-term memory was seen when the animals were tested 1.5 h after training, but not later. **d**, The effect of intra-entorhinal CNQX, MUS and NAN on short-term memory was seen when the animals were tested 1.5 or 3.0 h, but not 4.5 h, after training. In this last experiment, MUS was dissolved in the vehicle. $N=10$ for all groups. Asterisks, significant differences from both controls at $P<0.02$ in a two-tailed Mann-Whitney U -test. Bars are interquartile bars.

sessions for both experiments were as above, except that the foot-shock was omitted (so we were measuring how long it took for rats to step down to the grid, and we used this time as a measurement of their memory of the shock). We stopped measuring the time taken to step down to the grid after 180 seconds^{6–8}. This required the use of non-parametric statistics^{7,8}.

When given into the CA1 subregion of the hippocampus, CNQX and MUS impaired both short-term and long-term memory; NAN had no effect on either; and DPAT blocked short-term without affecting long-term memory (Fig. 1a). When given into the entorhinal cortex, CNQX, MUS and NAN blocked short-term memory without altering long-term memory, and DPAT enhanced short-term and blocked long-term memory (Fig. 1b). The effect of intrahippocampal DPAT on short-term

memory was no longer seen at 3 hours after treatment (Fig. 1c). The effect of intra-entorhinal CNQX, MUS and NAN on short-term memory lasted 3 hours (Fig. 1d).

We draw three conclusions. First, and foremost, four treatments (intrahippocampal DPAT and intra-entorhinal CNQX, MUS and NAN) block short-term memory without altering long-term memory. This shows for the first time, to our knowledge, that short-term and long-term memory mechanisms are separated. Second, the effects of the intra-entorhinal treatments on short-term memory lasted longer than that of intrahippocampal DPAT, suggesting that the intra-entorhinal receptors 'handle' short-term memory for a longer time. Third, similar receptors play different roles in the two brain structures at the same time during the period immediately after training. In the CA1 subregion, AMPA receptors are necessary for both short-term and long-term memory^{3,6}; GABA_A receptors inhibit both types of memory; and serotonin 1A receptors inhibit short-term but not long-term memory. In the entorhinal cortex, AMPA and serotonin 1A receptors are needed for short-term memory; serotonin 1A receptors inhibit long-term memory; and GABA_A receptors inhibit short-term memory.

We have shown previously that none of these treatments has any effect on short-term memory when given 0.1 hours before testing, or on long-term memory when given 1.5 hours before testing¹⁰. This rules out the possibility that our findings (Fig. 1) could be interpreted in terms of influences on retrieval or performance. The use of post-training infusions precluded the study of forms of short-term memory lasting less than 1.5 hours^{4,5}; drugs such as those used here take 1 hour to diffuse away from brain infusion sites¹¹.

So short-term and long-term memory involve essentially separate mechanisms. Some processes in the CA1 subregion (such as the involvement of AMPA receptors and the susceptibility to MUS) seem to be common to both memory systems, which should not be surprising as they both deal with nearly the same sensorimotor representations¹⁰.

Iván Izquierdo, Daniela M. Barros, Tadeu Mello e Souza, Marcia M. de Souza, Luciana A. Izquierdo

Departamento de Bioquímica, Instituto de Ciências Básicas da Saúde, Universidade Federal do Rio Grande do Sul, Ramiro Barcellos 2600, (90035-003) Porto Alegre, RS, Brazil

Jorge H. Medina

Instituto de Biología Celular 'Eduardo De Robertis', Facultad de Medicina, Universidad de Buenos Aires, Paraguay 2155 - 3er. Piso, (1121) Buenos Aires, Argentina

- James, W. *The Principles of Psychology* (Holt, New York, 1890).
- Gold, P.E. & McGaugh, J.L. in *Short-Term Memory* (eds Deutsch, D. & Deutsch, J.A.) 355–378 (Academic, New York, 1975).
- Izquierdo, I. & Medina, J.H. *Neurobiol. Learn. Mem.* **68**, 285–316 (1997).
- Squire, L.R., Knowlton, B. & Musen, G. *Annu. Rev. Psychol.* **44**, 453–495 (1993).
- Rosenzweig, M.R. et al. *Behav. Brain Res.* **57**, 193–198 (1993).
- Izquierdo, I. et al. *Eur. J. Neurosci.* **9**, 786–793 (1997).
- Bevilaqua, L. et al. *Behav. Pharmacol.* **8**, 331–338 (1997).
- Bernabeu, R. et al. *Proc. Natl Acad. Sci. USA* **94**, 7041–7046 (1997).
- Paxinos, G. & Watson, C. *The Rat Brain in Stereotaxic Coordinates* (Academic, San Diego, 1986).
- Izquierdo, I. et al. in *Monoamine Interactions and Brain Disease* (eds Palamo, T., Archer, R. & Beninger, R.) (Farrand, London, in the press).
- Martin, J.H. *Neurosci. Lett.* **127**, 160–164 (1991).

Walking on Mars

Sometime in the near future humans may walk in the reduced gravity of Mars. Gravity plays an essential role in walking. On Earth, the body uses gravity to 'fall forwards' at each step and then the forward speed is used to restore the initial height in a pendulum-like mechanism. When gravity is reduced, as on the Moon or Mars, the mechanism of walking must change¹. Here we investigate the mechanics of walking on Mars onboard an aircraft undergoing gravity-reducing flight profiles. The optimal walking speed on Mars will be 3.4 km h⁻¹ (down from 5.5 km h⁻¹ on Earth) and the work done per unit distance to move the centre of mass will be half that on Earth.

In contrast to swimming or flying, during which the fins or wings can slide against the surrounding medium, locomotion on a solid surface is constrained by the link between the centre of gravity of the body and the fixed point of contact of the foot on the ground. After foot contact, this link leads to a forward deceleration which must be compensated for by a subsequent forward acceleration in order to maintain a constant average speed of locomotion. This increases the cost of terrestrial locomotion.

However, this cost is contained by the transfer of kinetic energy into gravitational potential energy during the deceleration (when the body rides upwards on the leg after heel strike) and the subsequent recovery of kinetic energy from the potential energy during acceleration. The recovery of mechanical energy by this mechanism is described by

$$R = (W_f + W_v - W_{cm}) / (W_f + W_v)$$

where W_f is the work needed to increase the kinetic energy, W_v is the work to increase the potential energy, and W_{cm} is the work to increase the total mechanical energy of the centre of mass. With an ideal, frictionless pendulum, W_{cm} would be nil and R equal to

1. With walking on Earth, R attains a maximum of 0.7 at about 5.5 km h⁻¹ (Fig. 1) near the speed at which the energy expenditure per unit distance is at a minimum². At higher and lower speeds, R decreases and the energy cost increases.

We investigate the mechanics of locomotion on Mars in three male subjects (of weight and height, respectively, of 77 kg, 1.79 m; 92 kg, 1.93 m; 86 kg, 1.79 m) walking at different speeds in a Martian gravity (0.4g, maintained for about 30 s) on a force platform (3 m × 0.4 m) sensitive to the force exerted in both forward and vertical directions³. The platform was fixed to the floor of a KC-135 and an A300 Airbus aeroplane during the 23rd and 24th European Space Agency parabolic flight campaigns. The force signals from the plate (also measured in ref. 4) were analysed⁵ to determine W_{cm} and R .

Figure 1 shows that on Mars the maximum pendular recovery of mechanical energy R is reduced to 0.6 and occurs at the lower speed of 3.4 km h⁻¹. In spite of the lower R , the minimum mechanical work done per unit distance to maintain the motion of the centre of mass on Mars is about one half of that on Earth. In general, walking a given distance at any absolute speed will be cheaper on Mars than on Earth. In fact, the energy consumption measured during locomotion in simulated partial gravity is less than that at 1g (refs 6,7). A decrease in the maximum speed of walking was also observed in partial gravity simulators^{7,8}.

On Mars, then, both the optimal walk-

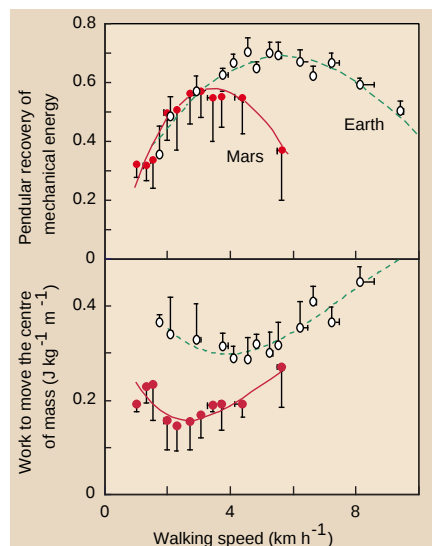


Figure 1 Comparison of the mechanics of walking on the Earth and on Mars. Top, mechanical energy recovered during walking by the pendular transfer between potential energy and kinetic energy of the centre of mass of the body on Mars and on Earth. Bottom, the work required to maintain the movement of the centre of mass in a sagittal plane. Note that the range of walking speeds on Mars is about half that on Earth (data on Earth were obtained from earlier studies on the same subjects^{9,10}). Mars, filled circles; Earth, open circles.

ing speed and the range of possible walking speeds will be about half those on Earth. The walk-run transition on Mars will occur near the optimal walking speed on Earth, and the mechanical work done to walk a given distance on Mars will be about half of what it would be on Earth. So, energy expenditure will probably be lower and locomotion smoother on Mars.

G. A. Cavagna

Istituto di Fisiologia Umana, Università di Milano, 20133 Milan, Italy

e-mail: cavagna@imiucca.csi.unimi.it

P. A. Willems, N. C. Heglund

Unité de Réadaptation, Université Catholique de Louvain, 1348 Louvain-la-Neuve, Belgium

- Margaria, R. & Cavagna, G. A. *Aerospace Med.* **35**, 1140–1146 (1964).
- Cavagna, G. A., Thys, H. & Zamboni, A. *J. Physiol., Lond.* **262**, 639–657 (1976).
- Heglund, N. C. *J. Exp. Biol.* **93**, 333–338 (1981).
- Rajulu, S. L., Klute, G. K. & Moore, N. R. *SAE Tech. Pap. Ser.* 22nd Intl Conf. on Environmental Systems, Seattle (1992).
- Cavagna, G. A. *J. Appl. Physiol.* **39**, 174–179 (1975).
- Farley, C. T. & McMahon, T. A. *J. Appl. Physiol.* **73**, 2709–2712 (1992).
- Newman, D. J., Alexander, H. L. & Webbon, B. W. *Aviat. Space Envir. Med.* **65**, 815–823 (1994).
- Hewes, D. E., Spady, A. A. Jr & Harris, R. L. *NASA Techn. Note TND-3363* (1966).
- Cavagna, G. A., Franzetti, P. & Fuchimoto, T. *J. Physiol., Lond.* **343**, 323–339 (1983).
- Willems, P. A., Cavagna, G. A. & Heglund, N. C. *J. Exp. Biol.* **198**, 379–393 (1995).

Relations of the new phylum Cycliophora

The Cycliophora is the most recently described animal phylum and is based on a single species, *Symbion pandora*, which was discovered on the mouthparts of the Norway lobster *Nephrops norvegicus*¹. Because only a few morphological data^{1–3} are available for *Symbion*, the precise nature of its phylogenetic relationships is highly controversial^{2,4,5}. Here we present a phylogenetic analysis of 18S ribosomal RNA sequence data, including a new *Symbion* sequence, which places *Symbion* in a lophophorate-aschelminth-protostome clade and which suggests a sister-group relationship between Cycliophora and a Rotifera-Acanthocephala clade.

N. norvegicus with attached *S. pandora* were collected in Kosterfjord, Sweden. We used the polymerase chain reaction to amplify 18S rRNA gene sequences from alcohol-preserved animals, and then used standard sequencing techniques to obtain 1,542 base pairs (bp) and 1,858 bp of the 18S rRNA gene sequences from *S. pandora* (EBI number Y14811) and *N. norvegicus* (Y14812), respectively. We added the sequences to an 18S rRNA database⁶. We used four tree-construction methods and

assessed the stability of the phylogenetic trees using bootstrapping (100 or 1,000 samples), confidence probability⁷ calculations and decay index⁸ calculations.

A neighbour-joining analysis of 18S rRNA sequences from a broad range of metazoan taxa placed *Symbion* in a lophophorate–aschelminth–protostome clade together with annelids, molluscs, rotifers, acanthocephalans, phoronids, brachiopods, ectoprocts, entoprocts, pogonophorans, vestimentiferans and echiurans (not shown). As this data set was too large to apply other reconstruction methods and stability tests, all further analyses were restricted to a data set of 36 taxa belonging to Arthropoda or to the lophophorate–aschelminth–protostome clade (Fig. 1). All reconstructions, for which the results are combined in a strict consensus tree (Fig. 1), show significant support for a sister-group relationship of Cycliophora to Rotifera and Acanthocephala. In agreement with previous findings^{9–11}, our analyses indicate that Acanthocephala is a subgroup of the Rotifera.

Our neighbour-joining analyses indicated that *Philodina acuticornis* and *Moliniformis moliniformis* are long branches, which may produce artificial groupings. Excluding both species did not affect the significantly supported nodes of Fig. 1 and resulted in only a small increase in the maximum parsimony consistency index (to 0.393). Therefore, it is unlikely that these long branches caused topological errors in our trees. Hence, the clustering of *P. acuticornis* and *M. moliniformis* is probably not an artefact.

The protostome affinities of Cycliophora are consistent with morphological data^{1,2}. *Symbion* has been tentatively proposed to be a sister group to Entoprocta. However, it also shares similarities with Ectoprocta and some aschelminth phyla². The 18S rRNA data do not provide evidence to suggest that cycliophorans may be neotenic entoprocts, but do indicate a relationship to Rotifera and Acanthocephala. Such a relationship was previously rejected^{1,2}, despite some morphological support. Both rotifers and cycliophorans have dwarf males and hypodermic insemination. As in sessile rotifers, the digestive tract of cycliophorans is U-shaped and the anus is located in the neck region outside the buccal funnel (the anus of sessile rotifers is located outside the corona). The digestive tract of entoprocts is U-shaped, too, but the anus is located inside the tentacular crown.

Moreover, the monophyly of the clade comprising Rotifera, Acanthocephala and Cycliophora is further supported by the common presence of protonephridia with multiciliated terminal cells. The plesiomorphic (or primitive) condition is one cilium per terminal cell, which is found in Gnathostomulida, Loricifera and some Priapulida. Kinorhyncha and several Priapulida have two cilia per cell. The apomorphic

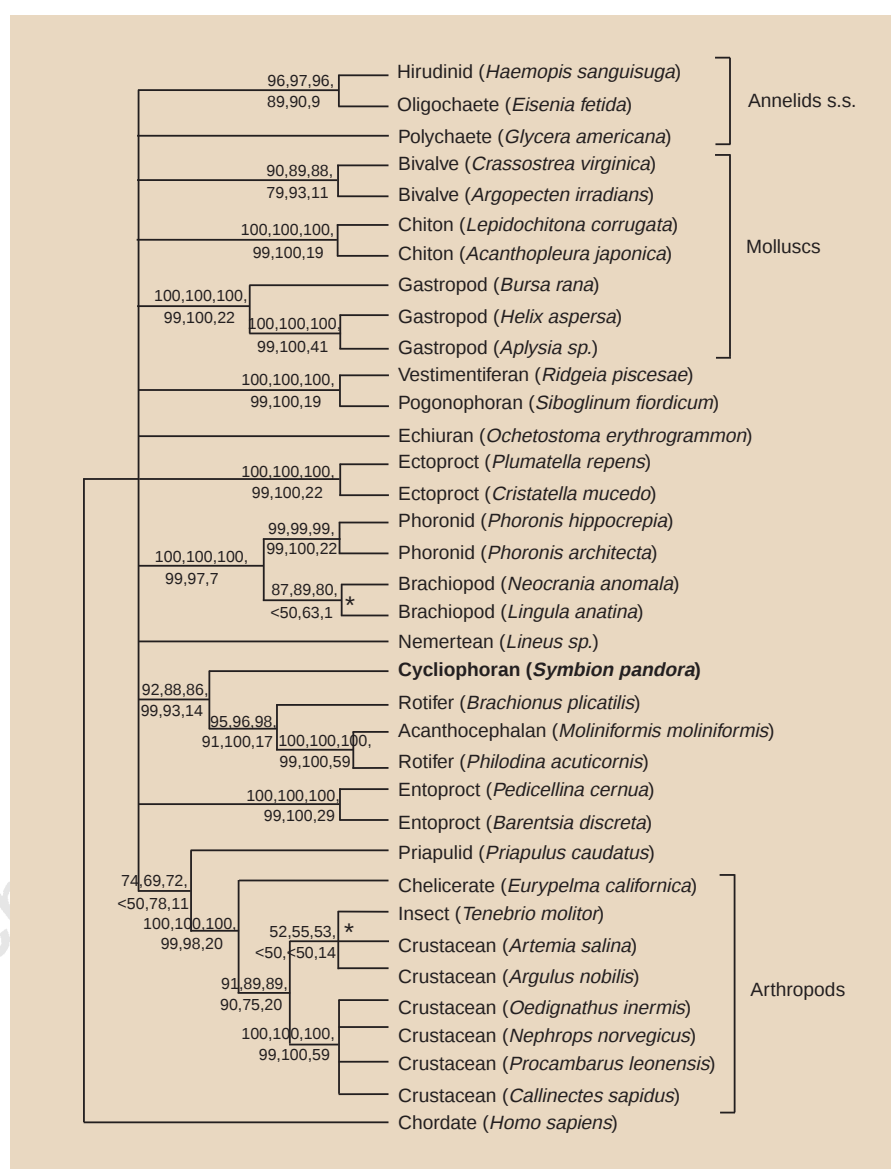


Figure 1 Strict consensus tree, showing groups present in all trees obtained by four different tree-construction methods applied to an 18S rRNA data set⁶ that includes the *S. pandora* and *N. norvegicus* data. Methods used were maximum parsimony analysis (two maximum parsimony trees, 3,850 steps, consistency index 0.383) and neighbour-joining analysis on the basis of Kimura¹³, Jukes-Cantor¹⁴ or substitution-rate-calibration¹⁵ distances. Numbers at nodes are, respectively, bootstrap values of these four methods, confidence probability values⁷ and decay indices⁸. All values are percentages, except for decay indices (number of extra steps for a clade not to be unequivocally supported). Nodes marked by asterisks were altered by removing long branches *P. acuticornis* and *M. moliniformis*.

(or derived) condition is the presence of many cilia per terminal cell as in Rotifera, Acanthocephala and now Cycliophora. Protonephridia are lacking in Ectoprocta, but are present in Entoprocta. Finally, some rotifers (for example, *Seison* and *Embata*¹²) have also adopted commensalistic lifestyles on crustacean gills.

Although the Cycliophora have several striking similarities with sessile rotifers, there is no evidence that they should be included as a subtaxon in Rotifera. The separate status of Cycliophora is supported by differences present in the cuticle and by the lack of a mastax, a feature always present in Rotifera. Both Acanthocephala and Rotifera have an

intracellular matrix layer and the epidermal cells consist of a syncytium, whereas *Symbion* has a true cuticle formed by normal epidermal cells. Owing to the basal position of *Symbion* in the Acanthocephala–Rotifera–Cycliophora clade, it is unlikely that the mastax has been secondarily lost in Cycliophora, as it has in the Acanthocephala.

B. M. H. Winnepeninckx, T. Backeljau
Royal Belgian Institute of Natural Sciences,
Vautierstraat 29, B-1000 Brussels, Belgium
e-mail: birgitta@uia.ua.ac.be

R. M. Kristensen
Zoological Museum, University of Copenhagen,
Universitetsparken 15, DK-2100 Copenhagen,
Denmark

1. Funch, P. & Kristensen, R. M. *Nature* **378**, 711–714 (1995).
2. Funch, P. & Kristensen, R. M. in *Microscopic Anatomy of Invertebrates 13: Lophophorates, Entoprocta and Cycliophora* (eds Harrison, F. W. & Woollacott, R. M.) 409–474 (Wiley-Liss, New York, 1997).
3. Funch, P. *J. Morphol.* **230**, 231–263 (1996).
4. Conway Morris, S. *Nature* **378**, 661–662 (1995).
5. d'Hondt, J.-L. *Bull. Mens. Soc. Linn. Lyon* **66**, 12–22 (1997).
6. Van de Peer, Y., Caers, A., De Rijk, P. & De Wachter, R. *Nucleic Acids Res.* **26**, 179–182 (1998).
7. Kumar, S., Tamura, K. & Nei, M. *MEGA: Molecular Evolutionary Genetics Analysis Version 1.0: Manual* (Pennsylvania State Univ., Pennsylvania, 1993).
8. Bremer, K. *Evolution* **42**, 795–803 (1988).
9. Lorenzen, S. in *The Origins and Relationships of Lower Invertebrates* (eds Conway Morris, S., George, J. D., Gibson, R. & Platt, H. M.) 210–223 (Clarendon, Oxford, 1985).
10. Garey, J. R., Near, T. J., Nonnemacher, M. R. & Nadler, S. A. *J. Mol. Evol.* **43**, 287–292 (1996).
11. Ahlrichs, W. H. *Zoomorphology* **117**, 41–48 (1997).
12. Brusca, R. C. & Brusca, G. J. *Invertebrates* (Sinauer, Sunderland, 1990).
13. Kimura, M. *J. Mol. Evol.* **16**, 111–120 (1980).
14. Jukes, T. H. & Cantor, C. R. in *Mammalian Protein Metabolism* (ed. Munro, H. N.) 21–132 (Academic, New York, 1969).
15. Van de Peer, Y., Van der Auwera, G. & De Wachter, R. *J. Mol. Evol.* **42**, 201–210 (1996).

Gulf Stream shifts following ENSO events

Over the past three decades the annual mean latitude of the Gulf Stream off the coast of the United States has been forecastable from the intensity of the North Atlantic Oscillation (NAO)¹, the predictions accounting for more than half the variance. Here we show that much of the unexplained variance can be accounted for by the Southern Oscillation in the Pacific, the Gulf Stream being displaced northwards following El Niño–Southern Oscillation (ENSO) events. This provides a link between events in the equatorial Pacific and the circulation and weather conditions of the North Atlantic.

Monthly charts showing the position of the north wall of the Gulf Stream since 1966 have been used to study its long-term variations^{2,3} and to construct an index of its latitude by principal components analysis¹. Ocean circulation theory predicts that the path of the Gulf Stream is set by the line of zero Ekman pumping, where there is no wind-driven convergence or divergence of water at the ocean surface⁴. The position of the Gulf Stream will therefore vary with the intensity of the North Atlantic Oscillation, the dominant mode of mid-latitude atmospheric variation in the North Atlantic, and indeed 60% of the annual variance in the latitude of the north wall 1966–96 is predictable from an NAO index^{1,5}.

Another source of fluctuation in the Gulf Stream is the subtropical and tropical trade wind belt. This is a region that is strongly affected by ENSO events in the Pacific Ocean, their influence appearing in African weather patterns⁶. Figure 1a shows that the residuals from a multiple regression prediction of the Gulf Stream position

based on the NAO index are correlated with the values from two years earlier of one measure of ENSO variations, the September-to-February averages of the Southern Oscillation (SO) index (the difference in sea-level pressure anomalies between Tahiti and Darwin). The SO index is uncorrelated with the NAO index⁷.

We used a stepwise regression analysis to relate the annual Gulf Stream position during 1966–97 to the previous year's position (to take account of persistence between years) and the NAO and SO indices at lags of up to two years. For each of these indices, the largest and most significant contribution was made by the value of the index from two years previously, there being no

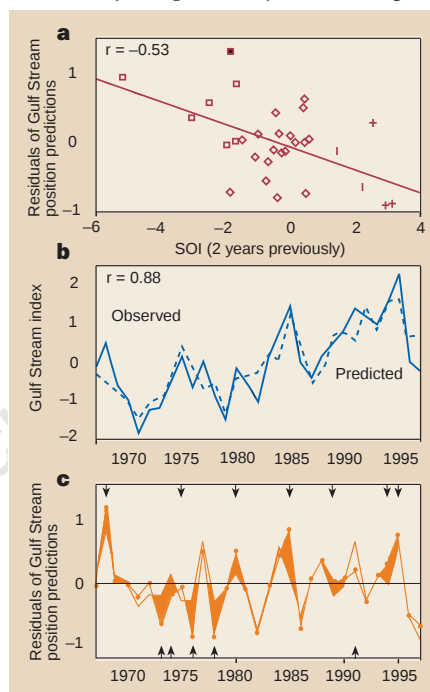


Figure 1 Observed and predicted Gulf Stream positions. **a**, Scatter plot comparing observed and predicted Gulf Stream positions (1966–97)¹ with the SO index two years previously (index has been allocated to the year of the January data). The prediction was based on a multiple regression equation using the NAO index⁵ and its value two years before, and the previous year's position. A regression line through the points is shown. Filled squares, the El Niño years (1966, 1973, 1978, 1983, 1987 and 1992–93); crosses, the La Niña years (1971–72, 1974, 1976 and 1989); diamonds, years belonging to neither group. **b**, Latitude of Gulf Stream 1966–97 (solid line) compared with the predictions of the regression equation (broken line) in which the delayed effect of the SO is added to the variables in **a**. The units of the index are equivalent to 0.03° at 79° W and 0.3° at 65° W. **c**, Reduction of regression residuals when delayed effect of the SO is considered (filled circles, original regression results). Downward arrows, two years after an El Niño; upward arrows, two years after a La Niña. Shaded regions, associated reductions of residuals. Correlation coefficients (r) both have $P < 0.01$. (All significance levels corrected for lag 1 serial correlation¹¹.)

significant contributions from the NAO and SO index at a one-year lag or from the unlagged SO index. The predictions of the regression equation (F -ratio = 22.7, $P < 0.01$) are shown in Fig. 1b. The two-year time lag is consistent with an earlier study of the latitude at which the Gulf Stream separates from the coast of the United States. Gangopadhyay *et al.*⁴ found that the theoretically predicted latitudes during 1977–88 only agreed with those observed if the wind forcing was integrated over about three years, a delay they attributed to the adjustment time of the ocean circulation.

Because of intercorrelation between the variables, it is not possible to obtain a unique value for the contribution of the SO index to the predictions in Fig. 1b. Thus the stepwise regression indicates that the index accounts for 9% of the variance, but the partial correlation coefficient between the Gulf Stream index and the SO index is 0.4, which indicates that the percentage could be larger. Figure 1c shows how including the lagged SO index reduces the regression residuals: after El Niño or La Niña events, the residual is reduced in the direction of the arrows.

Averaging the data from ref. 1 over the region from 65° to 79° W shows that the mean latitude of the Gulf Stream two years after ENSO events was 0.2° further north than after non-event years ($P < 0.001$). These displacements represent two-thirds of a standard deviation (the southward shift following La Niña events, 0.05°, was not significant). The displacements may be accompanied by larger shifts further east, where the Stream's path is less constrained, or by more substantial circulation changes elsewhere.

The position of the Gulf Stream affects the waters over the continental shelf from Cape Hatteras to the Grand Banks in several ways⁸ and may influence storm tracks over the northwest Atlantic^{9,10}. The results in Fig. 1 indicate that these processes may be linked to ocean–atmosphere events in the tropical Pacific.

Arnold H. Taylor, Michael B. Jordan, John A. Stephens

NERC Centre for Coastal and Marine Sciences, Plymouth Marine Laboratory, Prospect Place, West Hoe, Plymouth PL1 3DH, UK
e-mail: aht@pml.ac.uk

1. Taylor, A. H. & Stephens, J. A. *Tellus* **50A**, 134–142 (1998).
2. Miller, J. L. *J. Geophys. Res.* **99**, 5057–5064 (1994).
3. Drinkwater, K. F., Myers, R. A., Pettipas, R. G. & Wright, T. L. *Can. Data Rep. Hydrogr. Ocean Sci.* **125**, 1–103 (1994).
4. Gangopadhyay, A., Cornillon, P. & Watts, R. D. *J. Phys. Oceanogr.* **22**, 1286–1301 (1992).
5. Hurrell, J. W. *Science* **269**, 676–679 (1995).
6. Cane, M. A., Eshel, G. & Buckland, R. W. *Nature* **370**, 204–205 (1994).
7. Rogers, J. C. *Mon. Weath. Rev.* **112**, 1999–2015 (1984).
8. Myers, R. A. & Drinkwater, K. F. *J. Mar. Res.* **47**, 635–656 (1989).
9. Dickson, R. R. & Namias, J. *Mon. Weath. Rev.* **104**, 1255–1265 (1976).
10. Cione, J. J., Raman, S. & Pietrafesa, L. J. *Mon. Weath. Rev.* **121**, 421–430 (1993).
11. Taylor, A. H. *ICES J. Mar. Sci.* **52**, 711–721 (1995).

The origin of altruism

Unto Others: The Evolution and Psychology of Unselfish Behavior

by Elliott Sober and David Sloan Wilson
Harvard University Press: 1998. Pp. 394.
\$29.95, £19.95

John Maynard Smith

It is obvious that the parts of organisms have specific functions. Since Darwin, this has been explained by natural selection, but what is the target of selection? Should individual organisms be thought of as the units of selection, or must one also consider levels above or below the individual, such as groups of individuals, or genes within individuals? In particular, how can one account for altruistic behaviour, by animals and by humans, or supra-individual structures such as termite mounds or human institutions, which have features ensuring their survival?

These are the questions discussed in this book. Readers unfamiliar with evolutionary biology should be warned that these questions have been the topic of long and occasionally acrimonious debates, in which the authors have been active participants. Readers of this review should also be warned that I have also been a participant, often on the other side.

The first half of the book discusses the role of population structure in the evolution of animal behaviour, in particular of altruistic behaviour. Although I have been a participant, I find it quite hard to decide what the current debate is about. Is it about what the world is like, or about the best words to use when we describe it? When I first became interested, back in the 1960s, it seemed fairly clear. Although, most of the time, Darwin regarded the individual organism as the target of selection, it was quite common for biologists to speak as if organs or behaviours existed for the good of the species as a whole.

This view was expressed with particular clarity in Verne Wynne-Edwards' 1962 book *Animal Dispersion* (Oliver and Boyd), which argued that animals limit their breeding to prevent the population exceeding its food supply. His great merit was to see that, for this to be true, selection must be acting on the population as a whole: species whose members limit their breeding survive, whereas those that do not become extinct. It is not hard to see that this will not work, and that the observed behaviours — such as territorial behaviour — can be explained in other ways. Elliott Sober and David Sloan Wilson would agree, but they would also argue, I think correctly, that if a species is divided into partially isolated groups, each established by a few founders, this will affect the traits that evolve.

So, if the 'old' group selection no longer has supporters, and we agree that group structure is important, what are we arguing

about? I think the argument is largely semantic, and could not be settled by observation. Two examples will make this clearer.

First, consider Wilson's 'trait group' model, first published in 1975 but still very much part of his thinking. A population is divided into trait groups, and selection acts upon them. The members then disperse and mate randomly, and their offspring come together again in trait groups. There are two kinds of individual: altruists, who benefit (in terms of fitness) each of their fellow group members to a degree b , at a cost to themselves c ; and non-altruists, who do not. If costs and benefits combine additively, and groups are formed randomly, then altruism cannot evolve. But if altruists tend to associate with altruists, and non-altruists with non-altruists, then altruism can evolve. This conclusion is agreed. Sober and Wilson interpret this result as arising from a conflict between within-group selection (favouring non-altruists) and between-group selection (favouring altruists, if assortment is not random). But it can also be seen as an example of kin selection, an idea developed by William Hamilton in 1963: the values of b , c and degree of assortment required for altruism follow at once from his famous inequality, $b > rc$, where r is the

degree of relatedness. So we have a single model, but two ways of analysing it.

More briefly, here is a second example. Evolutionary game theory was first developed to explain the ritualistic nature of animal fights. This had often been explained in terms of the 'good of the species': it seemed desirable to George Price and myself to attempt an explanation in terms of individual selection. Sober and Wilson reconsider this model. They come to exactly the same conclusions that we reached, but argue that it is a case of group selection. Why? Essentially, because it is a model of the interaction of two individuals, and, for them, two individuals constitute a group. They ask: "is it really fair to call a pair of individuals a group, especially if they interact only briefly... never to meet again?" They conclude that it is. In effect, they say that any situation in which fitness is determined by interactions between individuals is a case of group selection.

Does it matter what words we use to describe a model if we agree about its consequences? Perhaps it does. We need formal models, but we also need intuition about why the models give the results they do, and the words used guide our intuitions and tell us what to look for. A group selection



Floral prints

In 1985, a group of botanists from the University of Oxford studied the flora of the Seychelles. Their artist, Rosemary Wise, spent the next ten years painting these unusual and threatened

plants. Her work, including *Nepenthes pervillei* (above), can be found in *A Fragile Eden: Portraits of the Endemic Flowering Plants of the Granitic Seychelles* (Princeton University Press, \$75).

terminology leads us to look for factors causing a difference between variation within and between groups; a kin selection model leads us to look for relatedness; and a game theory model leads us to look for frequency dependence and non-additive fitness interactions.

If this is right, the argument is not about what the world is like, or even about how we should model it (that is, what simplifying assumptions are adequate to explain it), but about what words we should use to explain our model. Is the trait group model an example of group selection or kin selection? Is the game theory model an example of group selection or individual selection?

For several reasons, I think the authors' approach to these questions is confusing. First, they are misleading for historical reasons. For example, when Price and I proposed our model of animal fighting, we were combating the then prevalent idea, supported by no less figures than Julian Huxley and Konrad Lorenz, that ritualized fighting behaviour had evolved for the good of the species. We tried to explain it by individual selection, not between-species selection. Essentially, we were arguing about what is an appropriate model of the world. It is therefore confusing to accept our model, but rename it group selection.

A second reason why the book is confusing is that, although the authors argue for pluralism, they are not themselves pluralists: for them, the only right way to describe a model is in group selection language. Any attempt to calculate the behaviour of a model by estimating the fitness of individuals is condemned as "fallacious averaging" (although Wilson himself carried out just such an averaging process in his original trait group model). It is "fallacious" because it detracts attention from the role of group-level processes (which, for them, include all interactions between individuals). The result of this bias is that it gives the impres-

sion the authors think that more than semantics is at issue.

Finally, and perhaps most important, they seem to confuse semantic and empirical issues. There are important empirical issues. For example, fascinating experiments by Michael Wade, Charles Goodnight and others show that selection between groups (and even between two-species communities) can be more effective than individual selection in producing change. Sober and Wilson describe these experiments, although they seem curiously uninterested in the underlying mechanisms. In discussing Wade's results they say only, in a footnote, "the reasons... are too technical to be treated in this book".

There follow two chapters concerned with the evolution of human societies. The emphasis is on the competition between human groups, and the role of social norms in guiding the behaviour of individual members of such groups. Because social norms can homogenize the behaviour of groups, the result is that such groups evolve properties ensuring group survival, just as individuals evolve traits ensuring individual survival. I found these chapters the most rewarding section of the book. The essential point is that higher-level entities (for example, individuals carrying many genes, or societies comprising many individuals) will evolve characteristics favouring the success of the group, provided there are processes that reduce within-group selection. In human groups, the most important such process is the homogenization of behaviour by social norms.

The final part of the book discusses human altruism and the motives that cause it. Do people help others because they think they will get pleasure from doing so (hedonism), or because they have an ultimate desire to help another (true altruism)? Sober and Wilson argue that evolutionary biology can shed light on this problem. They do not say that human traits that evolved by individual selection are hedonistic, and those that evolved by group selection are truly altruistic. Their argument is more subtle than that. They start by reviewing psychological and philosophical attacks on the question, and decide it has not been decisively answered. Their approach is to take a specific example, parental care, and argue that it increases biological fitness, and that the most effective proximate mechanism for generating parental care is altruism. It is a kind of reverse engineering argument. Such arguments are used in biology, but usually as a first step in a physiological investigation. For example, an effective proximate mechanism for bird migration would be a magnetic sense: this has led people to seek direct evidence for such a sense. It is a lot harder to seek direct evidence for mechanisms underlying true altruism. I am not really competent to judge this final part of the book; indeed, I am not sure I understood the distinction between

hedonistic and altruistic motives for helping others in the first place.

This book should carry a health warning. Read critically, it will stimulate thought about important questions. Swallowed whole, its effects would be disastrous. □

John Maynard Smith is in the School of Biological Sciences, University of Sussex, Falmer, Brighton BN1 9QG, UK.

New edition

A second edition of John Maynard Smith's *Evolutionary Genetics* has recently been published by Oxford University Press (£50, \$95, hbk; £19.95, \$39.95, pbk). When the first edition was reviewed in *Nature* 339, 107 (1989), R. C. Lewontin wrote: "When... a specialized subject is treated by an expert... the changes in the intellectual structure, the problems and the methods of the field can be seen. John Maynard Smith's new book is a model example."

Per ardua ad Stockholm

I Wish I'd Made You Angry Earlier: Essays on Science, Scientists, and Humanity

by Max Perutz

Cold Spring Harbor Laboratory Press: 1998. Pp. 354. \$39

Walter Gratzer

Speaking at a memorial symposium for A. Michelson, Einstein related the following anecdote: he had, he said, once asked Michelson why, when the velocity of light was already known with adequate precision, he was continuing to measure it with ever greater accuracy. "Because," Michelson had replied, "I get so much fun out of it." "Das finde ich wunderbar," was Einstein's verdict.

The parallel to Max Perutz's 60-year devotion to the red protein, haemoglobin, is not exact of course, for the deeper he has dug into its workings, the more rewarding the lessons that have emerged, and his work stands now as one of the pinnacles of this century's achievements. The pleasure he has drawn from his quest shines through this life-affirming selection of his writings, and he quotes with approval Noël Coward's dictum: "Work is fun. There is no fun like work."

The centrepiece of this collection is Perutz's account of his experiences during the Second World War, when, in the company of thousands of desperate refugees from the horrors unfolding in Germany, and of Italian chefs and waiters who had lived peaceably in Britain for decades, he was incarcerated on the Isle of Man and then deported to Canada. The ageing blimp who commanded the camp was heard to remark that he had had no idea so many of these

Science book prize

Jared Diamond, professor of physiology at the University of California Los Angeles School of Medicine, has won the 1998 Rhône-Poulenc science book prize for *Guns, Germs and Steel: The Fates of Human Societies* (Norton/Vintage; for a review see *Nature* 386, 339; 1997). This makes him the only author to have won the prize twice (his previous success was *The Third Chimpanzee*). What's more, *Guns, Germs and Steel* was also awarded this year's Pulitzer prize for general non-fiction (see *Nature* 392, 750; 1998). Other shortlisted authors for the £10,000 Rhône-Poulenc prize, which was announced last week at the Science Museum in London, were David Deutsch (*The Fabric of Reality*; Allen Lane, 1997), Richard Fortey (*Life*; Knopf), Ernst Mayr (*This Is Biology*; Belknap, 1997) and Simon Singh (*Fermat's Last Theorem*; Walker, 1997).



Max Perutz: extraordinary historical grasp.

Nazis were Jews. A ship, the infamous *Aran-dora Star*, was torpedoed, and most of the internees, confined below decks, were drowned. Perutz relates the harrowing experience of one Italian survivor, later professor of Italian at the University of Cambridge, who told it to him.

Perutz, eager to play his part in the war, was eventually rescued through the intervention of J. D. Bernal, and was drawn into a comically ill-starred project, conceived by the eccentric inventor Geoffrey Pike, to construct floating airstrips of reinforced ice in the Atlantic Ocean. He eventually returned to Cambridge and his haemoglobin crystals.

Perutz's favourite scientists, vividly portrayed in his pages with many good anecdotes, include the two great masters of his craft: the crystallographers W. L. Bragg, his admired mentor who started it all, and Dorothy Hodgkin. Another friend is François Jacob, whose outlook, civilized and humane, as his writings reveal, is so reminiscent of Perutz's own.

Perutz's benevolence deserts him only in the face of injustice, pomposity or casuistry. Here for instance is how he deftly slips the knife into Richard Lewontin, who proffers the oracular assertion that "Darwin's theory of evolution by natural selection is obviously [*sic*] nineteenth-century capitalism writ large, and his immersion in the social relations of a rising bourgeoisie had an overwhelming effect on the contents of his theory": "Marxism may be discredited in Eastern Europe," Perutz observes, "but it still seems to flourish at Harvard". He is infuriated by the school of science history which holds that all truth is relative and conditioned by the culture of time and place, for there is nothing

fugitive about the structure of DNA or haemoglobin.

Perhaps it was his impatience with the rhetoric of soft disciplines and his insistence on intellectual rigour that drew Perutz into the most unequivocal of scientific pursuits. The crystallographers established an ascendancy over chemists from the outset. When the Braggs solved the crystal structure of sodium chloride, Henry Armstrong, a prominent and highly vocal chemist, erupted in *Nature* in a spluttering paroxysm of affronted professional pride. His invective, which bore the title "Poor common salt", pronounced that no one with an ounce of chemical education could have put forward such a monstrous absurdity as a molecule that was not a molecule. Bragg himself told of how, at the end of a lecture, a chemist approached him to ask whether he might not see his way to moving the sodium atoms in the crystal even the tiniest bit closer to one or other chlorine.

Perutz's demolition of a revisionist biography of Pasteur, which derogates the man and his achievements, is magisterial (although it should be said that in the subsequent exchanges in the *New York Review of Books* he did not have it entirely his own way); and his analysis of a meretricious account of nuclear waste management in Britain is crushingly informed and incisive. In a sympathetic review of Ruth Sime's biography of Lise Meitner (whom he befriended in her old age in Cambridge), he nails the feminist canard that Otto Hahn deliberately sought to diminish her contribution to the discovery of nuclear fission and magnify his own.

Perutz also records the injustice done to Albert Schatz, who was robbed by Selman Waksman (to whom a misplaced Nobel prize eventually accrued) of the credit for discovering actinomycin. He is a mite hard perhaps on Andrei Sakharov for willingly working on the Soviet hydrogen bomb, and is not altogether correct in the assertion that Sakharov and his colleagues lied when they reported that they had achieved fusion in their first hydrogen-bomb test. Richard Rhodes in his definitive account, *Dark Sun*, states that 15–20% of the total yield came from fusion, compared with 24% from the first American thermonuclear device, exploded eight months earlier, with an admittedly much bigger bang.

A trifling distortion also occurs in the review of Hans Krebs's memoirs: the epic paper by Krebs on the citric acid cycle was not exactly rejected by the editor of *Nature*, Sir Richard Gregory, relying too confidently on his own scientific judgement. Rather, the editors of *Nature* presented their compliments to Krebs (as was their wont in those more courtly times) and begged to inform him that the Letters section of the journal was for the present fully supplied, but he was welcome to try his luck again a few months later.

Perutz remarks that, for all Krebs's loyalty to Britain, which had given him shelter in a

time of need, he remained ponderously Prussian in manner. Few Englishmen, he thinks, could have made such a sententious show of their upright principles as Krebs does in his memoirs. When in 1949 Perutz's student Francis Crick demolished a cherished conjecture about the structure of haemoglobin, he [Crick] would scarcely have solemnly recorded that he had acted "honestly, in good faith and in a spirit of helpfulness" (Krebs's words), rather than, as Perutz puts it, "admitting to a certain mischievous satisfaction in the dastardly deed". I can certainly confirm that, like the Bible, Krebs's book is unsullied by any trace of humour from first to last. And when Krebs asserts that he was not in the least discomposed by a false rumour that he had won a Nobel prize, Perutz recalls his own feelings at a time when similar rumours began to circulate in his laboratory. One day, two telegrams were delivered to the lab, one for John Kendrew and one for himself. But they turned out to be an enquiry about how many reprints they would require of a paper read at a recent conference. "We also," he confesses, "pretended to a stoic calm".

Many of these pieces were first published in the *New York Review of Books*, but I enjoyed them even more the second time than the first. Perutz's extraordinary historical grasp and the breadth of his personal experience and cultural perspective give his reviews an interest that often transcends that of the books themselves. He brings luminously to life such figures as Fritz Haber, Lise Meitner and Leo Szilard. He writes with wonderfully lucid precision about science and offers also a fine polemic, first read to the American Philosophical Society, on the meaning of freedom. The copy-editors might have done a better job of checking names, and I wonder whether the opera-loving Peter Medawar could really have alluded to Verdi's *Faust*? This, though, is a wholly captivating book; it has warmth, wit and style, and not a dull sentence. I urge you to read, enjoy and learn. □

Walter Gratzer is at the Randall Institute, King's College London, 26–29 Drury Lane, London WC2B 5RL, UK.

Chemistry gallery

Molecules at an Exhibition: Portraits of Intriguing Materials in Everyday Life

by John Emsley

Oxford University Press: 1998. Pp. 250.

£18.99, \$25

István Hargittai

Marcel Berthelot once pointed out that chemistry resembles the arts. It is unique among the natural sciences in that it creates most of its objects by synthesis. The exhibits in this book include both natural and

man-made substances, selected for being important, either useful or harmful, in our everyday life. Some, such as nitric oxide, have made a remarkable transition in our judgement. Eight galleries group the substances, from foodstuffs to vital components of our body, from illicit drugs to raw materials and energy sources, from agents that destroy us and our environment to those that give us pleasure.

Chemistry books, even those written for the general public, are often burdened with complex formulae. Fortunately, here John Emsley provides only a few familiar formulae, such as H_2O and CH_4 . Knowing the formulae of the more complex molecules he describes would not make it easier to understand their function.

A case in point is his description of the way we excrete unwanted nitrogenous material from the body by molybdenum-containing xanthine oxidase, a mammalian enzyme, producing uric acid. If uric acid is overproduced it accumulates in the form of sharp crystals in the joints, resulting in the painful illness of gout. The mechanism is understandable as described, and including the relevant formulae would not make it any easier to grasp the idea.

Most of us are largely unaware of the roles of many of the exhibited molecules in our everyday life. For example, phenylethylamine makes us feel good when we eat chocolate, sodium azide explodes on the impact of a car crash to save our lives, and thallium will poison us if the food of cows whose milk we drink had been exposed to thallium sulphate, which is used to kill rats.

Engaging stories are sprinkled throughout the galleries. Emsley tells us about the British discovery of penicillin and why British citizens had to pay US companies to use it. He explains that every time a young man thinks of sex — four times an hour, on average — his thoughts generate nitric oxide to help him fulfil his desires. And he provides a balanced assessment of the uses and dangers of DDT, one of the most worshipped and most feared molecules.

The descriptions are accurate without being pedantic, and the captivating short stories didactic without appearing patronizing. But this is a chemistry book nevertheless, albeit an unusual one in that it provides a lot of natural and cultural history along with its chemistry. A broad audience, regardless of whether it has a background in chemistry, will enjoy browsing and reading it. The composer Modest Mussorgsky's "Pictures at an Exhibition" is a classic; Emsley's own 'exhibition' will also be receiving many visitors for a long time to come. □

*István Hargittai is at the Budapest Technical University, St Gellert 4, Budapest H-1521, Hungary.
e-mail: hargittai.aak@chem.bme.hu*

In retrospect chosen by David Jones

The Scientist Speculates: An Anthology of Partly Baked Ideas

edited by Irving John Good, Alan James Mayne and John Maynard Smith (1962)

This book was published in 1962 from submissions solicited, received and edited during the previous year. The resulting compilation of 123 separate items was remarkably challenging and readable. Still more remarkable, it remains challenging and readable nearly 40 years later. My own copy continues to accumulate the scuffs and stains of use.

For a start, it is a triumph of the editor's art. Good and his fellow editors managed to interest some of the greatest names of the time in their project, including J. D. Bernal, Sir Cyril Burt, Dennis Gabor, J. E. Littlewood, N. W. Pirie, Sir Robert Robinson and Eugene E. Wigner (although how they kept J. B. S. Haldane out of it remains a mystery). Furthermore, they elicited from their contributors a splendid range of really high-class partly baked ideas. How many duds they rejected I cannot guess, but their selection was masterly.

The 'bakedness' of each individual idea may have risen or fallen since 1962, but few have grown stale. My guess is that, even now, a thoughtful scientist will find something of interest in at least 30% of the entries (although each reader will pick a different 30%).

An objective appraisal is clearly impossible; even a selection from my own favourites has to be pretty arbitrary. Donald Michie has a very perceptive essay on the difference between puzzles (in which one player confronts a physical challenge) and games (in which two or more players confront each other). The two require quite different skills. An animal experimenter who presents a rat with a puzzle, such as a maze, may scorn it for learning slowly. The rat, however, knows that life is a game, and against wily opponents. It is risky to stick to a single strategy, especially if it has succeeded once or twice. (It occurs to me that scientists are natural puzzle people; accordingly, they are everywhere outranked and outwitted by administrators and politicians, who are game people.)

Bernal remarks that water vapour rises from the sea, not because it is hot, but because its molecular weight is less than that of air. He suggests that large light chimneys and enclosures of plastics sheeting could entrain, condense and recycle water vapour for agricultural purposes. O. G. Selfridge has another large-scale application for plastics sheeting: submerging it in the coastal Pacific Ocean, thus warming up the surface water by the El Niño effect and dispersing the temperature inversion over Los Angeles.

Gabor discusses the dimensions of consciousness, generalizing from that small region of subjective experience (colour vision) that has already been mapped in three

dimensions. Pirie suggests the breeding of multipurpose plants, of which the roots, stem, leaves and seeds are all useful. W. H. Cazaly proposes that the instinct for cruelty in human beings arose from the need to enforce tribal discipline by punishment. It is quite distinct from sexual sadism. Marvin Minsky asks: "Where is our microtechnology?", and points out how fast tiny machines could work. These days, the fabrication of simple micromechanisms in silicon is raising the bakedness of this idea towards unity.

Even those ideas whose bakedness has shrunk since 1962 are still interesting. Good guesses, entirely reasonably, that computer technology should attain full artificial intelligence by 1978, at a cost of \$1 billion. What went wrong? A. D. Maude takes Alfvén's theory that sunspots are the surface manifestations of internal magnetohydrodynamic 'whirl rings' in the Sun, and develops a brilliant argument that they evolved as pseudo-living entities. It is wrong — or at least incompatible with modern sunspot theory — but might work in another context. S. C. Wallwork has a way of testing clairvoyance by guessing the signs of the structure factors of an X-ray crystallographic map. A set of guesses even slightly better than chance should start the map refining to the correct crystal structure. These days, computers do it all for us; but could clairvoyance help to sort out protein folding?

The practical realities of scientific life are also represented. They have changed little since 1962. C. D. Graham's "Technical glossary" (of which some entries are mis-ascribed to William McClimont) contains such gems as "Three of the samples were chosen for detailed study", which translates as "The results on the others didn't make sense and were ignored". J. de Bloggins (whom I suspect is Good himself in disguise) pays homage to the unholy trinity of Dirt, Noise and Leaks. He notes that data usually have at least 1% gross errors; as do people. Consoling, he points out that the only practical problem is what to do next.

The style and atmosphere of the book is consistently pleasing. For its contributors, speculation is a pleasure as well as a duty. They know their subject intimately, but are not bounded by its minutiae; none of their speculations is a mere anticipation of some narrow technical advance. They are prepared to play with ideas in any field, following them wherever they lead, looking all the time for ways of testing the results or checking them against theory. I fear that their vision of science is increasingly rare among the specialized, grant-bound, careerist researchers of today. A modern editor who attempted to bring out a second volume would find the first a hard act to follow. □

David Jones is in the Department of Chemistry, University of Newcastle, Newcastle upon Tyne NE1 7RU, UK.

Generation of hydrothermal megaplumes by cooling of pillow basalts at mid-ocean ridges

M. R. Palmer & G. G. J. Ernst

Department of Earth Sciences, University of Bristol, Bristol BS8 1RJ, UK

Hydrothermal megaplumes are huge volumes of anomalously warm water that are located up to 1,000 metres above the sea floor and appear to be generated at mid-ocean ridges. Since their discovery in 1986, there has been considerable debate concerning their origin. A theoretical model is used to argue that the cooling of pillow basalts, which are erupted at $\sim 1,200^\circ\text{C}$ into sea water and are the most common form of submarine volcanic activity, is responsible for the megaplume formation.

High-temperature hydrothermal activity is a spectacular manifestation of the formation of new crust at undersea spreading centres. Hot water issues from the sea floor at over 350°C and rises into the overlying water column, mixing with cold ambient sea water as it does so. Eventually, the rising plume entrains sufficient sea water to reach a level of neutral buoyancy where it then spreads laterally and is advected away from the vent site by deep ocean currents. The height that the plume rises above the sea floor is largely a function of the density stratification of the overlying sea water and the heat flux of the hydrothermal source. Plumes overlying vents of the 'black smoker' type typically rise 100–300 m above the sea floor and are emitted from sources that operate over periods of years; hence they are termed 'chronic plumes'. Recently, transient, detached, hydrothermal plumes have been observed up to 1,000 m above the sea floor. These features are termed megaplumes as they are generated over a period of days by a heat flux several orders of magnitude greater than that which generates chronic plumes (see ref. 1 for a review of chronic and transient hydrothermal plumes). Generation of megaplumes has been ascribed to rapid emptying of a high-temperature ($\sim 350^\circ\text{C}$) hydrothermal reservoir in the crust, of the type underlying black smokers^{2–5}, or the result of hydrothermal circulation within the oceanic crust initiated by dyke injection⁶.

The eruption of pillow basalts onto the sea floor is a ubiquitous feature of undersea spreading centres, and must involve thermal perturbation of the overlying water column as they are erupted at temperatures of $\sim 1,200^\circ\text{C}$ and cool to ambient conditions ($\sim 2^\circ\text{C}$) within days to weeks. We have examined the physical and chemical features that might be expected from cooling of pillow basalts in the deep oceans, and conclude that this process is a viable mechanism for the formation of hydrothermal megaplumes. This simple explanation has allowed us to constrain more closely the global flux and generation of megaplumes.

Previous observations and models

The first observation of a megaplume (called MPI) was over the southern Juan de Fuca ridge (SJFR)⁷. MPI had a thickness of 700 m and a diameter of 20 km. Its centre was ~ 700 m above the sea floor (depth of sea floor, 2,300 m) and its top reached $\sim 1,000$ m above the sea floor (as delineated by the 0.12°C temperature difference anomaly contour). The maximum temperature anomaly was only 0.28°C , but the large volume of the megaplume indicated that it contained a heat anomaly of 67×10^{15} J. Water samples from the plume contained large (5–130 μm) grains of anhydrite, chalcopyrite and amorphous silica. As the settling velocities of these particles

are ~ 70 – 200 m d^{-1} and the rise time of the plume to its neutral buoyancy height was ~ 1 h (ref. 8), the plume source must have been active (and forming hydrothermal anhydrite grains) only a few days before the plume was discovered. Baker *et al.*⁷ argued that the megaplume formed either by sudden release of hydrothermal fluid from underlying oceanic crust, or by rapid cooling of a surface of hot rock.

Baker *et al.*² detected another detached megaplume (called MPII) slightly to the north of the original sighting. From a model of plume formation from a line source, they recalculated the heat anomaly contained within MPI (120×10^{15} J) and presented the first calculation of the heat anomaly within MPII (58×10^{15} J). They concluded that the megaplumes formed over 2–20 d from a fissure 2,000–200 m long. Subsequently, additional detached megaplumes have been identified in the North Fiji basin⁹, the Eastern Manus basin¹⁰ and further north on the Juan de Fuca ridge¹¹.

The hypothesis that megaplumes are derived from cooling of a lava flow has generally been discarded^{2,3} on the grounds that; (1) plumes generated from a lava flow would not coalesce to form a single plume (that is, they would form a number of discrete plumes rising less far above the sea floor), (2) seawater interaction with a lava flow could not generate the chemical signature observed in megaplumes, and (3) pillow lavas show fresh surfaces that are incompatible with the extensive hydrothermal alteration required to produce megaplume chemical anomalies. However, Butterfield *et al.*¹² have recently suggested that a portion of the chemical anomalies observed in megaplumes over the 'CoAxial' segment of the Juan de Fuca ridge were due to high-temperature reaction between sea water and pillow basalts, although they did not consider the physical aspects of the problem.

Evidence that the SJFR megaplumes were related to tectonic and magmatic events is provided by their close temporal and spatial coincidence with the eruption of $\sim 50 \times 10^6$ m^3 of pillow lavas at the site¹³. The young pillow lavas were free of pelagic sediment, but were covered by varying amounts of hydrothermal sediment. From comparisons with subaerial systems, Chadwick and Embley¹³ estimated that the lava flow formed in a period of days to weeks. Submersible observations indicated that the lava flow was very permeable, and that there was evidence that heated water had flowed through the mounds of pillows and interacted chemically with the hot rock¹³; this was consistent with observations of pillow basalts from a drill core and ophiolites¹⁴. Cooling of $\sim 50 \times 10^6$ m^3 of pillow lavas to ambient temperature would liberate $\sim 215 \times 10^{15}$ J of heat, so the key questions to answer in assessing whether cooling of pillow lavas could generate a megaplume are; first, do the lavas cool

sufficiently rapidly to generate the heat flow required to generate merging turbulent plumes and achieve the rise height and heat anomaly of the megaplume, and second, is reaction between circulating sea water and hot pillow basalt sufficient to create the observed chemical anomalies?

Here we will show that; (1) theoretical considerations indicate that cooling of pillow basalts can yield the heat flux necessary to drive turbulent convection in sea water and generate megaplumes, (2) interactions with a high water/rock ratio do produce sufficient chemical fluxes, and megaplumes can also acquire components by entrainment of local chronic plumes, and (3) the degree of basalt alteration required to generate the chemical signature of megaplumes is low, and most of this alteration takes place within permeable pillow mounds that are not normally visible.

Heat flow

In considering the cooling of pillow basalts we have adopted the approach of Höskuldsson and Sparks¹⁵ (Fig. 1). The model assumes that pillow basalts are incrementally erupted onto the sea floor from a dyke and undergo conductive cooling. This heat source leads to convective heat flow in the overlying water column, creating plumes which merge into a single turbulent hydrothermal plume. We assume that a single layer of pillows is erupted over an area of sea floor (Fig. 1a) and then calculate the heat loss from that layer for a particular eruption rate. The initial layer of pillows is then buried by further monolayers (Fig. 1b), with the heat flow calculation repeated. The model adopts a conservative assumption that the heat loss from the pillows drops to negligible values after they are buried, although it is likely that the interaction of sea water with the buried pillow basalt will increase (Fig. 1b).

When basalts are erupted into sea water they undergo conductive cooling to yield a heat flux;

$$q_{\text{cond}} = \frac{k(T_m - T_c)}{\lambda \sqrt{\kappa_b \Delta t}}$$

where q_{cond} is the heat flux, k is the thermal conductivity of basalt ($2.9 \text{ W m}^{-1} \text{ K}^{-1}$), t is time, κ_b is the thermal diffusivity ($8.3 \times 10^{-7} \text{ m}^2 \text{ s}^{-1}$), T_m is the magma temperature ($1,200^\circ \text{C}$), T_c is the temperature of water in contact with the lava flow, Δt is the elapsed time and λ is a parameter of the order of unity obtained from the equation $L\pi^{1/2}/c(T_m - T_c) = \exp(-\lambda^2)/(\lambda \text{erf } \lambda)$, where L is the latent heat of crystallization ($2.09 \times 10^5 \text{ J kg}^{-1}$) and c is the heat capacity ($1.2 \times 10^3 \text{ J kg}^{-1} \text{ K}^{-1}$) (refs 16–18). The variation in heat flow per unit area with time for basalt magma cooling from $1,200$ to 0°C is shown in Fig. 2. The propagation distance of the temperature change $T_m - T_c$ into the cooling basalt is given by $(\kappa_b \Delta t)^{1/2}$ (ref. 17), hence basalt pillows with typical radii of 0.5 – 1 m (ref. 19) will undergo cooling in $(75\text{--}300) \times 10^3 \text{ s}$. These should be taken as maximum times, as the presence of prismatic water-cooled fractures (which typically have dimensions of 5 – 10 cm) in the pillows will accelerate cooling rates^{14,15}.

The recent pillow mound identified by Embley *et al.*²⁰ on the northern Cleft segment exemplifies the process of heat loss from cooling pillow lavas. The size of this mound is $\sim 1.25 \times 0.5 \text{ km}$. It is assumed that individual pillows were erupted to cover this area to produce a single layer of pillows, and that this layer was subsequently covered by a new layer. For typical pillow-basalt eruption rates of 10 – $1,000 \text{ m}^3 \text{ s}^{-1}$ (ref. 21) the mound would be covered every $(0.625\text{--}62.5) \times 10^3 \text{ s}$ for 1-m -diameter pillows, which is similar to, or less than, the time taken for the pillows to cool. Hence the average heat flux can be calculated from the time taken to form the new layer¹⁵. Examples of these calculations are shown in Fig. 2 for pillows of diameter 0.5 m and 1 m . Eruption rates of $100 \text{ m}^3 \text{ s}^{-1}$ would generate the SJFR lava flow¹³ in $\sim 6 \text{ d}$; this is consistent with experimental and theoretical studies²² (for an apparent basalt viscosity of 10^3 Pa s or higher). This time interval is also similar

to, but slightly shorter than, the time the model predicts for formation of the heat anomaly contained within MPI (9.9 d for 0.5-m pillows and 13.9 d for 1-m pillows). This small discrepancy may be accounted for by the conservative nature of the model assumptions; that is, that the individual pillows are only cooled conductively and that they cease to contribute heat to the megaplume after they are buried. Nevertheless, the model demonstrates that the pillow cooling rates are compatible with the timescales and magnitude of hydrothermal heat flux required for megaplume formation.

It is likely that there will be convection of heated sea water along the margins of the dyke feeding the pillow lava flow. Indeed, Lowell and Germanovich⁶ suggested that hydrothermal circulation along the margins of a dyke of dimensions 10 km strike length, extending 1 km deep into the crust, and 1 m wide could generate a megaplume without cooling of associated pillow basalts. We recognize that this mechanism could contribute to the heat flux required to generate a megaplume, but we believe that it is unlikely to fuel a megaplume on its own. Cooling of a dyke can be approached in the same manner as above, although the surface area of an equivalent volume of magma in a dyke is considerably less than that contained within pillow basalts. The surface area to volume ratio for 0.5-m and 1-m diameter pillows is 12 and 6 m^{-1} , respectively, compared to 2 m^{-1} for a 1-m -wide dyke. Hence, under the conditions described above, a lava eruption rate of $100 \text{ m}^3 \text{ s}^{-1}$ would form MPI in 10 – 14 d for cooling of 0.5 – 1 m pillows, but would require at least 24 d for cooling of a dyke.

This calculation assumes that conductive cooling of the dyke is

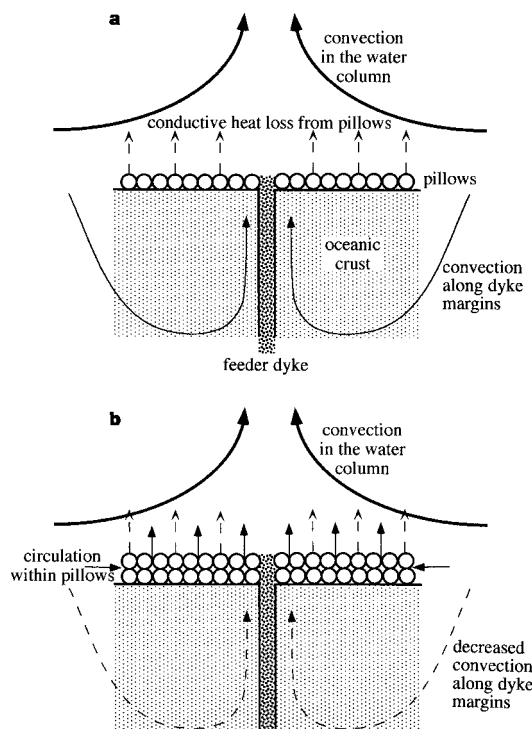


Figure 1 Cartoon illustrating assumptions used in model calculations. **a**, A monolayer of pillow lava ('pillows') is erupted over an area of sea floor, and the heat loss from that layer is calculated for a particular eruption rate and pillow size. **b**, The initial layer of pillows is buried by further monolayers and the heat flow calculation repeated. It is likely that the convective heat flow along the sides of the feeder dyke will decrease as the overlying layer of pillow basalts thickens. In addition, the degree of water-rock reaction within the pile of pillow basalts will increase as the number of layers increases.

the limiting factor (rather than the porosity and permeability of the oceanic crust) for heat transport to the overlying water column, and that the sea water cooling the dyke has a temperature of 0 °C. In reality the temperature of the sea water will increase as it rises up the dyke walls and is progressively heated, thus reducing the conductive cooling rate of the dyke. At higher temperatures hydrothermal fluids precipitate secondary minerals at an increasing rate, reducing the permeability of the crust adjacent to the dyke and further reducing the heat flux. In addition, dykes do not terminate at the sea floor, but generate overlying sheet flows and/or pillow basalts which form a cap over the dyke and may further decrease the rate of cooling of the dyke. Another problem is that magma flow would become focused within hours to days for a 10-km-long, 1-m-wide linear source, and is likely to lead to discrete plumes rising kilometres apart from each other, rather than coalescing to form a single plume rising from a roughly circular mound of pillow basalts. However, as sea water convecting along the dyke walls will reach higher temperatures than that cooling pillow basalts, it is likely to have higher chemical/heat ratios and may contribute significantly to the chemical anomalies within the megaplume.

Turbulent convection and plume generation

In a submarine eruption heat is extracted from the cooling lava to drive convection in the overlying sea water. Convective heat transfer will occur as long as the heated sea water remains less dense than unheated background sea water and as long as the Rayleigh number, based on the lava flow width, remains $> 2 \times 10^3$ (ref. 23). In practice, this implies that when the heated sea water has a tempera-

ture of a few degrees higher than that of ambient, convection will cease and thereafter heating of the overlying water will be conductive. The calculations presented below suggest that this transition occurs after ~ 23 d; this is consistent with the estimate of Baker *et al.*² that the maximum duration of discharge of the source for MPI was ~ 20 d.

During the convective heat-transfer phase, conservation of heat requires that the heat flux out of the lava by conduction (q_{cond}) equals the convective heat flow (q_{conv}) in the heated sea water. This convective heat flow is given by¹⁵;

$$q_{\text{conv}} = \rho_w c_w J (T_c - T_w)^{4/3}$$

where ρ_w and c_w are the density and heat capacity of sea water, respectively, T_w is the ambient water temperature (~ 0 °C) and J is given by;

$$J = \gamma \left(\frac{\alpha g \kappa_l^2}{\nu} \right)^{1/3}$$

where γ is an empirical constant whose value is close to 0.1, α is the liquid expansion coefficient, ν is the liquid kinematic viscosity, κ_l is the liquid thermal diffusivity and g is the acceleration due to gravity^{2, 22}. We note that α , ν and κ_l have been derived from steam tables at the appropriate data for sea water are not available. However, we do not believe that any small uncertainties that might result will invalidate the basic conclusions of our model. Plotted in Fig. 2 are values of q_{conv} for selected ΔT values (the temperature difference between the heated and ambient sea water) between 10 °C and 80 °C. Figure 2 shows that the cooling lava can heat the water and result in natural convection in the overlying water column, provided that the pillows release most of their heat on timescales of $< 10^4 - 10^5$ s.

After a given pillow unit has been emplaced (see Fig. 2) and a time of the order of 10^5 s has elapsed, $\Delta T_{\text{conv}} (= T_c - T_w)$ will be of the order of (20 °C) and the Rayleigh number of the convective flow above an equidimensional lava flow of 500 m width will be of the order of 10^{19} , that is, well within the fully turbulent regime. Dimensional analysis suggests that the velocity of the ascending turbulent flow (U) immediately above the lava flow scales as $U \sim (g \alpha \Delta T)^{1/2}$ (where D is the typical length scale of the Rayleigh number; ref. 23, p. 173), that is, of the order of 3 m s^{-1} . Using this value in the expression defining the Rossby number, $Ro \sim U/\Omega D$ (where Ω is the angular velocity), one finds $Ro \gg 1$ for typical values of $\Omega = 7 \times 10^{-5} \text{ rad s}^{-1}$ and a lava flow of the specified dimensions. This limit of negligible rotation, together with laboratory studies²⁴, implies that individual plumes forming above each pillow in the lava flow will typically merge after a few metres of rise above the source to form a single plume. Generation of a cyclonic flow would be expected around the rising plume column. The azimuthal velocities (u_t) associated with cyclonic flow scale as $u_t \sim (F_o f)^{1/4}$, where $F_o \sim Ag(\Delta\rho/\rho)(g\alpha L\Delta T)^{1/2}$ (where A is the lava flow area) is the source buoyancy flux and f is the Coriolis frequency²⁵; that is, for typical values of $t \approx 10^4 - 10^5$ s, $F_o \approx (50 - 300) \times 10^3 \text{ m}^4 \text{ s}^{-3}$, $f = 10^{-4} \text{ s}^{-1}$, u_t values will be up to approximately $1 - 2 \text{ m s}^{-1}$. Above a $1,250 \times 500 \text{ m}$ pillow lava flow, Ro is of the order of 1 after about 10^6 s (~ 20 d) and the plume is expected to take the form of an intense vortex flow²⁴ that is cyclonic toroidal.

Far above the source, the plume flow will be dominated by an anticyclonic vortex, as previously modelled for point sources²⁵ and observed in laboratory experiments²⁴. This geostrophic vortex is controlled by gravity and the Earth's rotation, and would typically require a venting duration of longer than 1 d to form. The timescale of 10–14 d for generation of the megaplume (Fig. 2) is in agreement with independent estimates; these estimates are based on experiments of the dependence on the Earth's rotation of the instability of megaplume generation²⁶.

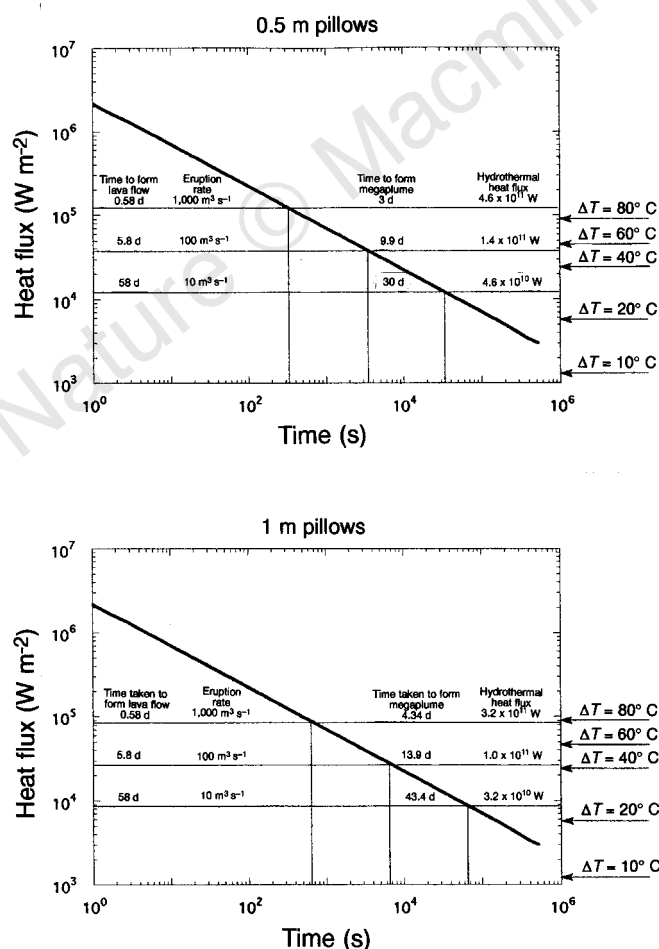


Figure 2 The variation of heat flux with time for pillows of diameter 0.5 m and 1 m (see text for details).

Chemical fluxes

The Fe and Mn concentrations of megaplumes and vent fluids on the SJFR are well characterized²⁷. These show that the ratios of Fe and Mn to heat (Q) in megaplumes are significantly lower than those in local vent fluids (Table 1). The lower Fe/ Q ratios may be partly due to sedimentation of Fe-rich particulate phases, but dissolved Mn is conserved over the lifetimes of the plumes sampled by Massoth *et al.*²⁷. This suggests that the chemical and heat anomaly in the megaplumes were not simply derived from the fluid reservoir that fuels local chronic hydrothermal activity.

Hydrothermal alteration of pillow basalts is seawater-dominated compared to the systems responsible for formation of black smokers, leading to suggestions that cooling of a lava flow could not generate the chemical anomalies in megaplumes^{2,3}. However, in water-dominated basalt–seawater experiments (water/rock = 125, $T = 300^\circ\text{C}$) Seyfried and Mottl²⁸ observed a rapid increase in dissolved Fe and Mn concentrations in the first sample taken (after 5 d) due to a drop in pH resulting from incorporation of $\text{Mg}(\text{OH})_2$ into secondary minerals. The Mn/ Q and Fe/ Q ratios in these experimental fluids (0.16×10^{-9} and $1.7 \times 10^{-9} \text{ MJ}^{-1}$, respectively) are similar to the values observed in megaplumes. A rapid increase in dissolved Fe and Mn levels was also seen in experiments (water/rock = 3, $T = 360^\circ\text{C}$) by Rosenbauer and Bischoff²⁹. In this case, dissolved Fe levels reached close to maximum values within 3 h (when the first sample was taken). After 10 h dissolved Fe concentrations fell until they reached $\sim 10\%$ of their maximum values after ~ 300 h. Dissolved Mn increased more slowly than Fe, but peaked after 10 h and, thereafter, declined by a factor ~ 3 . These experiments indicate that higher dissolved Fe and Mn concentrations are expected in a kinetically controlled hydrothermal system, compared to one in which equilibrium conditions are reached, that is, the high Fe/Mn ratios in the megaplumes are consistent with generation by a seawater-dominated hydrothermal system.

Basalts on the SJFR contain 0.20 wt% ($29 \times 10^{-6} \text{ Mg}^{-1}$) MnO and 11.4 wt% ($1.5 \times 10^{-3} \text{ Mg}^{-1}$) FeO (ref. 20), yielding total Mn and Fe contents in the recent pillow lava flow of 4.3×10^9 and $230 \times 10^9 \text{ M}$, respectively. Hence, the Mn and Fe in MPI could have been generated by leaching of 0.32% of the total Mn and 0.036% of the total Fe within the pillow-basalt lava flow. The extraction efficiencies of Mn and Fe (relative to Li) for hydrothermal fluids from the SJFR are 0.072 and 0.0056, respectively³⁰. If these efficiencies are applied to hydrothermal leaching of the pillow basalts, this implies that the degree of alteration of the pillows required to produce the Mn and Fe in MPI are 4.4% and 6.5%, respectively. However, the behaviour of Mn and Fe in the experiments discussed above^{28,29} suggests that the required degree of alteration may be closer to 0.5%. It is likely that most of the hydrothermal alteration of the pillow basalts will take place within the mounds, which are at least 45 m thick¹³. Hence, surface observations describing the relatively pristine nature of the pillow basalts¹¹ are unlikely to provide a reliable record of their extent of alteration. In addition, Butterfield *et al.*¹² observed the presence of fresh crystals of sulphide minerals and halite on the surfaces of fresh pillow basalts that were indicative of high-temperature reaction between the basalt and sea

water. These minerals are unstable in sea water, so that their absence on other submarine basalts cannot be taken to indicate that pillow basalts do not undergo hydrothermal alteration.

The largest hydrothermal particles identified in MPI were composed of anhydrite²⁷. Anhydrite is generated when sea water alone is heated above 150°C (ref. 28), hence its presence in megaplumes does not require any water–rock interaction. Models of megaplume generation all assume that the plume formed by entrainment of ambient bottom water with physical characteristics identical to those of bottom water outside the vent area^{2–4}. As the SJFR megaplumes formed in an area of chronic hydrothermal venting, it is likely that entrainment drew in material from these systems. Numerical models of episodic discharge⁸ suggest that entrainment into the rising MPI would result in radial current velocities of 0.05 m s^{-1} extending for $>1 \text{ km}$ from the megaplume generation site and up to 300 m above the sea floor. The exact site of MPI and MPII generation on the SJFR is uncertain as they were detached from their sources when they were detected and there are deep water currents in the area². Nevertheless, the presence of black smokers within 1 km of the fresh pillow basalts¹³ suggests that entrainment of portions of the chronic hydrothermal plumes by actively forming megaplumes would have been likely, regardless of how the latter were formed. Chronic hydrothermal activity releases $(0.6\text{--}5.0) \times 10^9 \text{ W}$ in this area². If all this heat flux was entrained into MPI over 10 d (the approximate duration of the MPI generation²) it would only account for about 4% of the heat anomaly. However, the relatively high Mn/ Q and Fe/ Q ratios in the vent fluids (Table 1) suggest they could account for up to 40% of the chemical anomaly in MPI if they were entrained into the rising megaplume. Hence entrainment of chronic hydrothermal plumes would reduce the extent of alteration of the pillow basalts required to generate the chemical anomalies in MPI.

Evidence from the CoAxial megaplumes

Following detection of a seismic event on the northern Juan de Fuca ridge, a cruise to the area discovered three detached megaplumes with heat anomalies of $(2.2\text{--}12.4) \times 10^{15} \text{ J}$ (ref. 11). SeaBeam (high-precision bathymetry) surveys revealed that the seismicity was accompanied by formation of at least $5 \times 10^6 \text{ m}^3$ of fresh pillow basalts³¹. The average Mn/ Q and Fe/ Q ratios of these megaplumes³² are similar to those from the SJFR (Table 1). However, there was no evidence for prior existence of chronic high-temperature hydrothermal activity in the CoAxial area. This led Lupton *et al.*³³ to suggest that the chemical and heat anomalies in the CoAxial megaplumes were generated by release of stagnant fluids resident in the crust, rather than by the release of fluids from a chronic hydrothermal system. We suggest that the similarity in style of the megaplume events at the SJFR and CoAxial sites indicates that they formed by the same process; that is, rapid cooling of a pillow basalt lava flow by sea water.

The global picture

In assessing the influence of megaplumes on ocean chemistry and circulation, a critical unknown is their frequency. Baker³⁴ was unable to answer this question directly, but noted that the large 1986 MPI probably occurred $\sim 30\text{--}1,000$ times less frequently than the smaller event plumes of 1987 and 1989. If, as we suggest, megaplumes result from extrusion of pillow basalts, we can set more precise limits on the extent of megaplume formation. For the purposes of illustration we take the Ocean Drilling Program (ODP) borehole 504B as typical of oceanic crust. Slightly more than the upper 1 km of the oceanic crust at this site is composed of extrusive pillow basalts³⁵. For a global oceanic crustal generation rate of $2.94 \text{ km}^2 \text{ yr}^{-1}$ (ref. 36), this corresponds to a global volume of pillow basalts of about $3 \text{ km}^3 \text{ yr}^{-1}$. If we assume that the heat anomaly contained in MPI ($120 \times 10^{15} \text{ J}$) was generated by extrusion of $50 \times 10^6 \text{ m}^3$ of pillow basalts (see above) and apply this same

Table 1 Chemistry/heat ratios of vent fluids and megaplumes

	Average Mn/ Q (nM J^{-1})	Average Fe/ Q (nM J^{-1})
Vent fluids		
South Cleft	2.64	12.3
North Cleft	1.25	3.47
Megaplumes		
MPI	0.11	0.84
MPII	0.08	0.13
CoAxial	0.14	0.72

ratio to the global oceans, then about 60 megaplumes would be erupted annually and would be responsible for $7.2 \times 10^{18} \text{ J yr}^{-1}$ ($230 \times 10^9 \text{ W}$) of heat loss from the oceanic crust. This is about 7.5% of recent estimates of the total hydrothermal heat loss from oceanic crust that is 0–1 Myr old³⁷.

The potential influence of megaplumes on the deep circulation of present-day oceans has been considered^{8,25}, but there is also interest in the consequences of penetration of the thermocline by buoyant hydrothermal plumes³⁸. Such an event could have profound consequences for a variety of processes, such as the development of atmospheric hurricanes³⁹, if the plume penetration occurred at low latitudes³⁸. Speer³⁸ calculated that in order for a hydrothermal plume to penetrate the thermocline with the present density stratification of the low-latitude oceans, it would have to be a factor of $\sim 10^3$ stronger than the biggest megaplume observed to date. This would seem to require an unfeasibly large source, regardless of the manner in which megaplumes are formed. However, there is evidence that the density gradient of the oceans has varied with time, and that density gradients at lower latitudes were less steep during the Cretaceous period than those of today⁴⁰, and it has been suggested that upwelling megaplumes could have penetrated the thermocline during this period, with climatic and ecological consequences^{40,41}. By explicitly and quantitatively linking the generation of megaplumes to the eruption of pillow basalts, we expect to be able to quantify the extent and consequences of megaplume formation for the volcanic extrusion rates and water stratification present in Cretaceous oceans.

The wider radius of entrainment at the base of megaplumes and their greater rise height above the sea floor compared to chronic hydrothermal plumes means that they provide a powerful mechanism for the dispersal of the larvae of organisms endemic to hydrothermal vents^{42,43}. Megaplumes may be of particular importance in this regard on slow-spreading ridges. For example, on the Mid-Atlantic Ridge the rise height of plumes from the most powerful chronic hydrothermal system found to date is well below that of the median valley walls⁴⁴, confining this mechanism of larval dispersion to a rift segment. In contrast, the greater rise height of megaplumes would allow entrained larvae to be transported far more widely. □

Received 20 January; accepted 14 April 1998.

1. Baker, E. T., German, C. R. & Elderfield, H. in *Seafloor Hydrothermal Systems* (eds Humphris, S. E., Zierenberg, R. A., Mullineaux, L. S. & Thomson, R. E.) 47–71 (Geophys. Monogr. 91, Am. Geophys. Union, Washington DC, 1995).
2. Baker, E. T., Lavelle, J. W., Feely, R. A., Massoth, G. J. & Walker, S. L. Episodic venting of hydrothermal fluids from the Juan de Fuca Ridge. *J. Geophys. Res.* **94**, 9237–9250 (1989).
3. Cann, J. R. & Strens, M. R. Modelling periodic megaplume emission by black smoker systems. *J. Geophys. Res.* **94**, 12227–12237 (1989).
4. Cathles, L. M. A capless 350 °C flow zone model to explain megaplumes, salinity variations, and high temperature veins in ridge axis hydrothermal systems. *Econ. Geol.* **88**, 1977–1988 (1993).
5. Wilcock, W. S. D. A model for the formation of transient event plumes above mid-ocean ridge hydrothermal systems. *J. Geophys. Res.* **102**, 12109–12121 (1997).
6. Lowell, R. P. & Germanovich, L. N. Dike injection and the formation of megaplumes at ocean ridges. *Science* **267**, 1804–1807 (1995).
7. Baker, E. T., Massoth, G. J. & Feely, R. A. Cataclysmic hydrothermal venting on the Juan de Fuca Ridge. *Nature* **329**, 149–151 (1987).
8. Lavelle, J. W. & Baker, E. T. A numerical study of local convection in the benthic ocean induced by episodic hydrothermal discharges. *J. Geophys. Res.* **99**, 16065–16080 (1994).
9. Nojiri, Y., Ishibashi, J., Kawai, T., Otsuki, A. & Sakai, H. Hydrothermal plumes along the North Fiji Basin spreading axis. *Nature* **342**, 667–670 (1989).
10. Gamo, T. et al. Hydrothermal plumes in the eastern Manus Basin, Bismark Sea: CH₄, Mn, Al and pH. *Deep-Sea Res.* **40**, 2335–2349 (1993).

11. Baker, E. T. et al. Hydrothermal event plumes from the CoAxial seafloor eruption site, Juan de Fuca Ridge. *Geophys. Res. Lett.* **22**, 147–150 (1995).
12. Butterfield, D. A. et al. Seafloor eruptions and evolution of hydrothermal fluid chemistry. *Phil. Trans. R. Soc. Lond. A* **355**, 369–386 (1997).
13. Chadwick, W. W. & Embley, R. W. Lava flows from a mid-1980s submarine eruption on the Cleft segment, Juan de Fuca Ridge. *J. Geophys. Res.* **99**, 4761–4776 (1994).
14. Rosenberg, N. D., Spera, F. J. & Haymon, R. M. The relationship between flow and permeability field in seafloor hydrothermal systems. *Earth Planet. Sci. Lett.* **116**, 135–153 (1993).
15. Höskuldsson, A. & Sparks, R. S. J. Thermodynamics and fluid dynamics of effusive subglacial eruptions. *Bull. Volcanol.* **59**, 219–230 (1997).
16. Carslaw, H. S. & Jaeger, J. C. *Conduction of Heat in Solids* 2nd edn (Oxford Univ. Press, 1959).
17. Turcotte, D. L. & Schubert, G. *Geodynamics, Applications of Continuum Physics to Geological Problems* (Wiley, New York, 1982).
18. Huppert, H. E. & Sparks, R. S. J. Melting of a chamber containing a hot turbulently convecting fluid. *J. Fluid Mech.* **188**, 107–131 (1988).
19. Walker, G. P. L. Morphometric study of pillow-size spectrum among pillow lavas. *Bull. Volcanol.* **54**, 459–474 (1992).
20. Embley, R. W., Chadwick, W., Perfit, M. R. & Baker, E. T. Geology of the northern Cleft segment, Juan de Fuca Ridge: Recent lava flows, sea-floor spreading and the formation of megaplumes. *Geology* **19**, 771–775 (1991).
21. Thórarinnsson, S. On the rate of lava and tephra production and the upward migration of magma in four Icelandic eruptions. *Geol. Rundsch.* **57**, 705–718 (1967).
22. Griffiths, R. W. & Fink, J. H. Solidification and morphology of submarine lavas: a dependence on extrusion rate. *J. Geophys. Res.* **97**, 19729–19737 (1992).
23. Tritton, D. J. *Physical Fluid Mechanics* (Clarendon, Oxford, 1988).
24. Boubnov, B. M. & Van Heijst, G. J. F. Experiments on convection from a horizontal plate with and without background rotation. *Exp. Fluids* **16**, 155–164 (1994).
25. Helfrich, K. R. & Speer, K. G. in *Seafloor Hydrothermal Systems* (eds Humphris, S. E., Zierenberg, R. A., Mullineaux, L. S. & Thomson, R. E.) 347–356 (Geophys. Monogr. 91, Am. Geophys. Union, Washington DC, 1995).
26. Woods, A. W. & Bush, J. W. M. The dimensions and dynamics of megaplumes. *J. Geophys. Res.* (submitted).
27. Massoth, G. J. et al. Temporal and spatial variability of hydrothermal manganese and iron at Cleft segment, Juan de Fuca Ridge. *J. Geophys. Res.* **99**, 4905–4923 (1994).
28. Seyfried, W. E. & Mottl, M. J. Hydrothermal alteration of basalt by seawater under seawater-dominated conditions. *Geochim. Cosmochim. Acta* **46**, 985–1002 (1982).
29. Rosenbauer, R. J. & Bischoff, J. L. in *Hydrothermal Processes at Seafloor Spreading Centers* (eds Rona, P. A., Boström, K., Laubier, L. & Smith, K. L.) 177–197 (Plenum, New York, 1983).
30. Von Damm, K. L. & Bischoff, J. L. Chemistry of hydrothermal solutions from the Southern Juan de Fuca Ridge. *J. Geophys. Res.* **92**, 11334–11346 (1987).
31. Chadwick, W. W., Embley, R. W. & Fox, C. G. SeaBeam depth changes associated with recent lava flows, CoAxial segment, Juan de Fuca Ridge: Evidence for multiple eruptions between 1981–1993. *Geophys. Res. Lett.* **22**, 167–170 (1995).
32. Massoth, G. J. et al. Observations of manganese and iron at the CoAxial seafloor eruption site, Juan de Fuca Ridge. *Geophys. Res. Lett.* **22**, 151–154 (1995).
33. Lupton, J. E. et al. Variations in water column ³He/heat ratios associated with the 1993 CoAxial event, Juan de Fuca Ridge. *Geophys. Res. Lett.* **22**, 155–158 (1995).
34. Baker, E. T. A 6-year time series of hydrothermal plumes over the Cleft segment of the Juan de Fuca Ridge. *J. Geophys. Res.* **99**, 4889–4904 (1994).
35. Bratt, S. R. & Purdy, G. M. Structure and variability of oceanic crust on the flanks of the East Pacific Rise between 11° and 13° N. *J. Geophys. Res.* **89**, 6111–6125 (1984).
36. Williams, D. L. & Von Herzen, R. P. Heat loss from the Earth: new estimate. *Geology* **2**, 327–328 (1974).
37. Stein, C. A., Stein, S. & Pelayo, A. M. in *Seafloor Hydrothermal Systems* (eds Humphris, S. E., Zierenberg, R. A., Mullineaux, L. S. & Thomson, R. E.) 425–445 (Geophys. Monogr. 91, Am. Geophys. Union, Washington DC, 1995).
38. Speer, K. G. Thermocline penetration by buoyant plumes. *Phil. Trans. R. Soc. Lond. A* **355**, 443–458 (1995).
39. Emanuel, K. A., Speer, K., Rotunno, R., Srivastava, R. & Molina, M. Hypercanes: A possible link in global extinction scenarios. *J. Geophys. Res.* **100**, 13755–13765 (1995).
40. MacLeod, K. G. & Huber, B. T. Reorganization of deep ocean circulation accompanying a Late Cretaceous extinction event. *Nature* **380**, 422–425 (1996).
41. Vogt, R. Volcanogenic upwelling of anoxic, nutrient-rich water: a possible factor in carbonate-bank/reef demise and benthic faunal extinctions. *Geol. Soc. Am. Bull.* **101**, 1225–1245 (1989).
42. Kim, S. L., Mullineaux, L. S. & Helfrich, K. R. Larval dispersal via entrainment into hydrothermal vent plumes. *J. Geophys. Res.* **99**, 12655–12665 (1994).
43. Mullineaux, L. S. & France, S. C. in *Seafloor Hydrothermal Systems* (eds Humphris, S. E., Zierenberg, R. A., Mullineaux, L. S. & Thomson, R. E.) 408–424 (Geophys. Monogr. 91, Am. Geophys. Union, Washington DC, 1995).
44. Rudnicki, M. D. & Elderfield, H. Theory applied to the mid-Atlantic hydrothermal plumes: the finite difference approach. *J. Volcanol. Geotherm. Res.* **50**, 161–172 (1992).

Acknowledgements. We thank S. Sparks, R. James and A. Woods at Bristol for their help and advice, and K. Speer and B. Boubnov for comments on the manuscript. Hydrothermal research at Bristol is supported by NERC grants from the BRIDGE programme. G.G.J.E. also thanks the 'Foundation Belge de la Vocation' for previous support.

Correspondence and requests for materials should be addressed to M.R.P. (e-mail: m.r.palmer@bris.ac.uk).

Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody

Peter D. Kwong*, Richard Wyatt†, James Robinson‡, Raymond W. Sweet§, Joseph Sodroski†|| & Wayne A. Hendrickson*‡

* Department of Biochemistry and Molecular Biophysics, and ‡ Howard Hughes Medical Institute, Columbia University, New York, New York 10032, USA

† Department of Cancer Immunology and AIDS, Dana-Farber Cancer Institute, and ‡ Department of Pathology, Harvard Medical School, and || Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, Massachusetts 02115, USA

‡ Department of Pediatrics, Tulane University Medical Center, 1430 Tulane Avenue, New Orleans, Louisiana 70112, USA

§ Department of Immunology, SmithKline Beecham Pharmaceuticals, 709 Swedeland Road, King of Prussia, Pennsylvania 19406, USA

The entry of human immunodeficiency virus (HIV) into cells requires the sequential interaction of the viral exterior envelope glycoprotein, gp120, with the CD4 glycoprotein and a chemokine receptor on the cell surface. These interactions initiate a fusion of the viral and cellular membranes. Although gp120 can elicit virus-neutralizing antibodies, HIV eludes the immune system. We have solved the X-ray crystal structure at 2.5 Å resolution of an HIV-1 gp120 core complexed with a two-domain fragment of human CD4 and an antigen-binding fragment of a neutralizing antibody that blocks chemokine-receptor binding. The structure reveals a cavity-laden CD4-gp120 interface, a conserved binding site for the chemokine receptor, evidence for a conformational change upon CD4 binding, the nature of a CD4-induced antibody epitope, and specific mechanisms for immune evasion. Our results provide a framework for understanding the complex biology of HIV entry into cells and should guide efforts to intervene.

The human immunodeficiency viruses HIV-1 and HIV-2 and the related simian immunodeficiency viruses (SIV) cause the destruction of CD4⁺ lymphocytes in their respective hosts, resulting in the development of acquired immunodeficiency syndrome (AIDS)^{1,2}. The entry of HIV into host cells is mediated by the viral envelope glycoproteins, which are organized into oligomeric, probably trimeric spikes displayed on the surface of the virion. These envelope complexes are anchored in the viral membrane by the gp41 transmembrane envelope glycoprotein. The surface of the spike is composed primarily of the exterior envelope glycoprotein, gp120, associated by non-covalent interactions with each subunit of the trimeric gp41 glycoprotein complex^{3,4}. Comparison of the gp120 sequences of different primate immunodeficiency viruses identified five variable regions (V1–V5) (ref. 5). The first four variable regions form surface-exposed loops that contain disulphide bonds at their bases⁶. The conserved gp120 regions form discontinuous structures important for the interaction with the gp41 ectodomain and with the viral receptors on the target cell. Both conserved and variable gp120 regions are extensively glycosylated⁶. The variability and glycosylation of the gp120 surface probably modulate the immunogenicity and antigenicity of the gp120 glycoprotein, which is the main target for neutralizing antibodies elicited during natural infection⁷.

Entry of primate immunodeficiency viruses into the host cell involves the binding of the gp120 envelope glycoprotein to the CD4 glycoprotein, which serves as the primary receptor. The gp120 glycoprotein binds to the most amino-terminal of the four immunoglobulin-like domains of CD4. Structures of both the N-terminal two domains^{8,9} and the entire extracellular portion of CD4 (ref. 10) have been determined, and mutagenesis indicates that the CD4 structure analogous to the second complementarity-determining region (CDR2) of immunoglobulins is critical for gp120 binding^{11,12}. Conserved gp120 residues important for CD4 binding have likewise been identified by mutagenesis^{3,13,14}.

CD4 binding induces conformational changes in the gp120 glycoprotein, some of which involve the exposure and/or formation of a binding site for specific chemokine receptors. These chemokine receptors, mainly CCR5 and CXCR4 for HIV-1, serve as obligate second receptors for virus entry^{15,16}. The gp120 third variable (V3) loop is the principal determinant of chemokine receptor specificity¹⁷. However, other more conserved gp120 structures that are exposed upon engagement of CD4 also seem to be involved in chemokine-receptor binding. This CD4-induced exposure is indicated by the enhanced binding of several gp120 antibodies^{18,19} which, like V3-loop antibodies, efficiently block the binding of gp120–CD4 complexes to the chemokine receptor²⁰. These are called CD4-induced (CD4i) antibodies. CD4 binding may trigger additional conformational changes in the envelope glycoproteins. For example, binding of CD4 to the envelope glycoproteins of some HIV-1 isolates induces the release, or ‘shedding’, of gp120 from the complex²¹, although the relevance of this process to HIV entry is uncertain.

HIV and related retroviruses belong to a class of enveloped fusogenic viruses that include corona-, paramyxo- and orthomyxo-viruses (e.g. influenza virus), all of which require post-translational cleavage for activation. The transmembrane coat proteins of these viruses (gp41 equivalents) share sequence similarity, particularly in their N-terminal fusion peptides, and they participate directly in membrane fusion. The ectodomain of gp41 can form a coiled coil resembling that of influenza haemagglutinin HA₂ (refs 3, 4, 22), supporting the idea that this class of viruses may share some common features of virus entry. In other respects, enveloped viruses tend to be distinctive. They use varying modes of entry (direct membrane penetration for HIV, endocytosis for influenza virus) and even otherwise closely related viruses may use individualized receptors. The exterior coat proteins (gp120 equivalents) are accordingly specialized. Thus, for example, there is no detectable similarity in sequence, nor now in structure, between the receptor-

binding portion of HIV and that of murine leukaemia virus²³, another retrovirus. Mechanisms for receptor-mediated triggering of fusion may also be virus-specific.

Because of the important role of the gp120 glycoprotein in receptor binding and interactions with neutralizing antibodies, information about the gp120 structure is important for understanding HIV infection and for the design of therapeutic and prophylactic strategies. Here we report the crystal structure at 2.5 Å resolution of a partially deglycosylated HIV-1 gp120 core bound to a two-domain fragment of the CD4 cellular receptor and to the antigen-binding fragment (Fab) of an antibody, 17b, that is directed against a CD4i epitope. The accompanying Letter relates this structure to the antigenic properties of gp120 envelope proteins²⁴.

Structure determination

Because of the extensive glycosylation and conformational heterogeneity associated with the HIV gp120 glycoprotein, we devised a crystallization strategy aimed at radical modification of the protein surface. We made truncations at termini and variable loops in various combinations with gp120 from various strains, extensively deglycosylated these gp120 variants, and produced complexes with various ligands. A theoretical analysis indicated that the probability of crystal formation is greatly increased by such reduction of surface

heterogeneity and trials with multiple variants²⁵. After screening almost twenty combinations of gp120 variants and ligands, we obtained crystals²⁵ of a ternary complex composed of a truncated form of gp120, the N-terminal two domains (D1D2) of CD4, and a Fab from the human neutralizing monoclonal antibody 17b (ref. 18).

The gp120 crystallized was from the HXBc2 strain of HIV-1. It has deletions of 52 and 19 residues from the N and C termini, respectively; Gly-Ala-Gly tripeptide substitutions for 67 V1/V2 loop residues and 32 V3 loop residues; and the removal of all sugar groups beyond the linkages between the two core *N*-acetylglucosamine residues. This deglycosylated core gp120 is stripped of over 90% of the carbohydrate but it retains over 80% of the non-variable-loop protein. Its capacity to interact with CD4 and relevant antibodies is preserved at or near wild-type levels²⁶. The crystals are of space group *P*222₁ (*a* = 71.6, *b* = 88.1, *c* = 196.7 Å), with one ternary complex and 58% solvent in the crystallographic asymmetric unit.

The ternary structure was solved by a combination of molecular replacement, isomorphous replacement and density modification techniques. It has been refined to an *R*-value of 21.0% (5–2.5 Å data > 2σ, *R*-free = 30.3%). The final model, composed of 7,877 atoms, comprises residues 90–396 and 410–492 of gp120 (except loop substitutions), residues 1–181 of CD4, and residues 1–213 of the light chain and 1–229 of the heavy chain of the 17b monoclonal antibody. In addition, 11 *N*-acetylglucosamine and 4 fucose residues, and 602 water molecules have been placed. The overall structure of the complex of gp120 with D1D2 of CD4 and Fab 17b is as shown in Fig. 1.

Structure of gp120

The deglycosylated core of gp120 as dissected from the ternary complex approximates a prolate ellipsoid with dimensions of 50 × 50 × 25 Å, although its overall profile is more heart-shaped than circular. Its backbone structure is shown in Fig. 2a, c in an orientation precisely perpendicular to that in Fig. 1 (Fig. 4e shows a mutually perpendicular view). This core gp120 comprises 25 β-strands, 5 α-helices and 10 defined loop segments, all organized with the topology shown in Fig. 2b. Specific spans of structural elements are given in Fig. 2d. The structure confirms the chemically determined disulphide-bridge assignments⁶ (Fig. 2c). The polypeptide chain of gp120 is folded into two major domains, plus certain excursions that emanate from this body. The inner domain (inner with respect to the N and C termini) features a two-helix, two-strand bundle with a small five-stranded β-sandwich at its terminoproximal end and a projection at the distal end from which the V1/V2 stem emanates. The outer domain is a stacked double barrel that lies alongside the inner domain so that the outer barrel and inner bundle axes are approximately parallel.

The proximal barrel of the outer-domain stack is composed from a six-stranded, mixed-directional β-sheet that is twisted to embrace helix α2 as a seventh barrel stave. The distal barrel of the stack is a seven-stranded antiparallel β-barrel. The two barrels share one contiguous hydrophobic core, and the staves also continue from one barrel to the next except at the domain interface. This interruption is centred at a side between barrels where the chain enters the outer domain, with loop LB insinuated as a tongue between strands β16 and β23. The extended segment just preceding LB is like an eighth stave of the distal barrel, but is slightly out of reach for hydrogen bonding with its β16 and β19 neighbours. The chain returns to complete the inner domain after β24.

The proximal end of the outer domain includes variable loops V4 and V5 and loops LD and LE, which are variable in sequence as well. Loop LC is also at this end, close in space to loop LA of the inner domain, although by topology it is at the other end of this domain. The distal end does include the base of the excised variable loop V3 and also an excursion via loop LF into a β-hairpin, β20–β21, which



Figure 1 Overall structure. The ribbon diagram shows gp120 in red, the N-terminal two domains of CD4 in yellow, and the Fab 17b in light blue (light chain) and purple (heavy chain). The side chain of Phe 43 on CD4 is shown. The prominent CDR3 loop of the 17b heavy chain is evident in this orientation. Although the complete N and C termini of gp120 are missing, the positions of the gp120 termini are consistent with the proposal that gp41, and hence the viral membrane, is located towards the top of the diagram. This would position the target membrane at the diagram base. The vertical dimension of gp120 in this orientation is roughly 50 Å. Perpendicular views of gp120 are shown in Figs 2 and 4. Drawn with RIBBONS⁴⁹.

in turn hydrogen-bonds with the V1/V2 stem emanating from the inner domain. This completes an antiparallel, four-stranded 'bridging sheet' which stands as a peculiar minidomain in contact with, but distinct from, the inner and outer domains as well as the excised V1/V2 domain. This bridging sheet also participates in the spatially separated interactions of gp120 with both CD4 and the 17b antibody (Fig. 1, and see below). One further excursion from the body of the outer domain produces strand β 15 and helix α 3, which are also important in CD4 binding.

Taken as a whole, the structure of gp120 seen here has no precedent. Moreover, our domain-level searches have failed to reveal similarity of the inner domain to any known atomic structures, although the missing terminal segments might conceal relationships. We do, however, find fragmentary similarity for portions of the outer domain with known structures. In particular, part of the protomer of FabA dehydrase²⁷ is like part of the proximal barrel, and dUTP pyrophosphatase²⁸ has elements in common with

both barrels of the outer domain. In each case, the superimposable fraction is limited. For FabA, 45 of its 171 C α atoms superimpose on five segments, but the rest are topologically unrelated. For dUTPase, 41 of its 152 C α atoms appropriately capture 8 of the 15 segments in the outer domain body, but there is no helix corresponding to α 2 and the placements of termini are not comparable. Several viruses related to HIV encode dUTPases; however, we have not found sequence evidence to support a role in coat protein evolution.

This structure of core gp120 should be a prototype for the class. As shown in the structure-based alignment of representative sequences (Fig. 2d), there is substantial conservation despite the noted variability among HIV strains. Thus, even an HIV-2 sequence is 35% identical with that of the HXBc2 strain expressed in this crystallized construct, and the identity rises to 77 and 51%, respectively, for the more closely related HIV-1 clade C and clade O representatives. The inner domain is appreciably more conserved than the outer domain, with 86, 72 and 45% identity for the C, O

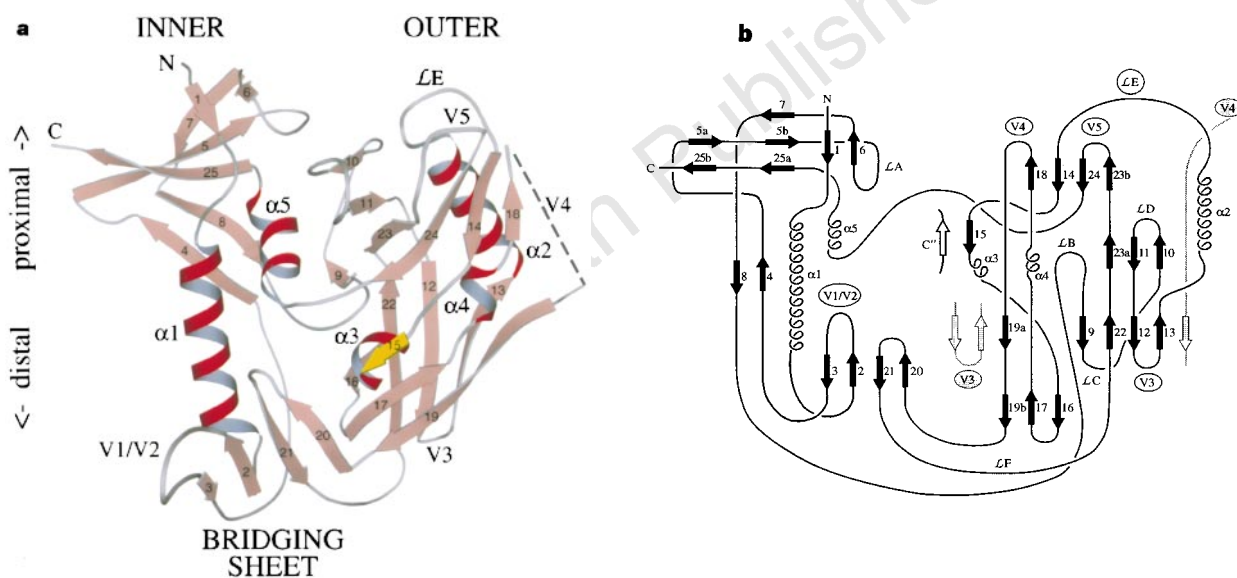


Figure 2 Structure of core gp120. In **a–c**, the orientation of gp120 is related to Fig. 1 by a 90° rotation about a vertical axis. Thus the viral membrane would be oriented above, the target membrane below, and the C-terminal tail of CD4 would be coming out of the page. In this view, we describe the left portion of core gp120 as the inner domain, the right portion as the outer domain, and the 4-stranded sheet at the bottom left of gp120 as the bridging sheet. The bridging sheet (β 3, β 2, β 21, β 20) can be seen packing primarily over the inner domain, although some surface residues of the outer domain, such as Phe382, reach in to form part of its hydrophobic core. **a**, Ribbon diagram. α -Helices are depicted in red and β -strands in salmon, except for strand β 15 (yellow), which makes an antiparallel β -sheet alignment with strand C' of CD4. Connections are shown in grey, except for the disordered V4 loop (dashed line) connecting β 18 and β 19. Selected parts of the structure are labelled. **b**, Topology diagram. The diagram is arranged to coincide with the orientation of **a**, **c**. Helices are shown as corkscrews and labelled α 1– α 5. β -Strands are shown as arrows: black and labelled represent the 25 β -strands of core gp120; grey and unlabelled represent the continuation of hydrogen bonding across a sheet; white and labelled represents the C' strand of CD4. Spatial proximity between neighbouring strands implies main-chain hydrogen bonding. Loops are labelled L A–L F and V1–V5. Labels for loops with high sequence variability are circled. Assignments of secondary structure were made with the Kabsch and Sander algorithm, except for β 4 and β 8, which are both interrupted mid-strand by side-chain-backbone hydrogen bonds, β 9, β 15 and β 25a, all of which have angles or hydrogen bonds that are slightly non-standard, and α 4, which hydrogen bonds as a 3_{10} helix, with the final residue in β conformation. **c**, Stereo plot of an α -carbon trace. Every 10th C α is marked with a filled circle, and

every 20th residue is labelled. Disulphide connections are depicted as ball and stick. The ordered residues 90–396 and 410–492 are shown. **d**, Structure-based sequence alignment. The sequences are shown of HIV-1 B (core gp120 from clade B, strain HXBc2 used in these studies), C (HIV-1 clade C, strain UG268A2), O (HIV-1 clade O, strain ANT70), HIV-2 (strain ROD), and SIV (African green monkey isolate, clone GRI-1). The secondary-structure assignments are shown as arrows and cylinders, with a cross denoting residues that are disordered in the present structure. The 'gars' sequence at the N terminus and the 'gag' sequence in the V1/V2 and V3 loops are consequences of the gp120 truncation. Solvent accessibility is indicated for each residue by an open circle if the fractional solvent accessibility is greater than 0.4, a half-filled circle if it is 0.1 to 0.4, and a filled circle if it is less than 0.1. Sequence variability among primate immunodeficiency viruses is indicated below the solvent accessibility by the number of horizontal hash marks: 1, residues conserved among all primate immunodeficiency viruses; 2, conserved among all HIV-1 isolates; 3, moderate variation among HIV-1 isolates; and 4, significant variability among HIV-1 isolates. In assessing conservation, all single atom changes were permitted as well as larger substitutions if the character of the side chain was conserved (for example, K to R or F to L). N-linked glycosylation is indicated by 'm' for the high-mannose additions and 'c' for the complex additions in mammalian cells⁶. Residues of gp120 in direct contact with CD4 are indicated by an asterisk. Direct contact is a more restrictive criterion of interaction than the often-used loss of solvent-accessible surface; residues of gp120 that have lost solvent-accessible surface but are not in direct contact include 123, 124, 126, 257, 278, 282, 364, 471, 475, 476 and 477. Panels **a** and **b** were drawn with MOLSCRIPT (P. J. Kraulis).

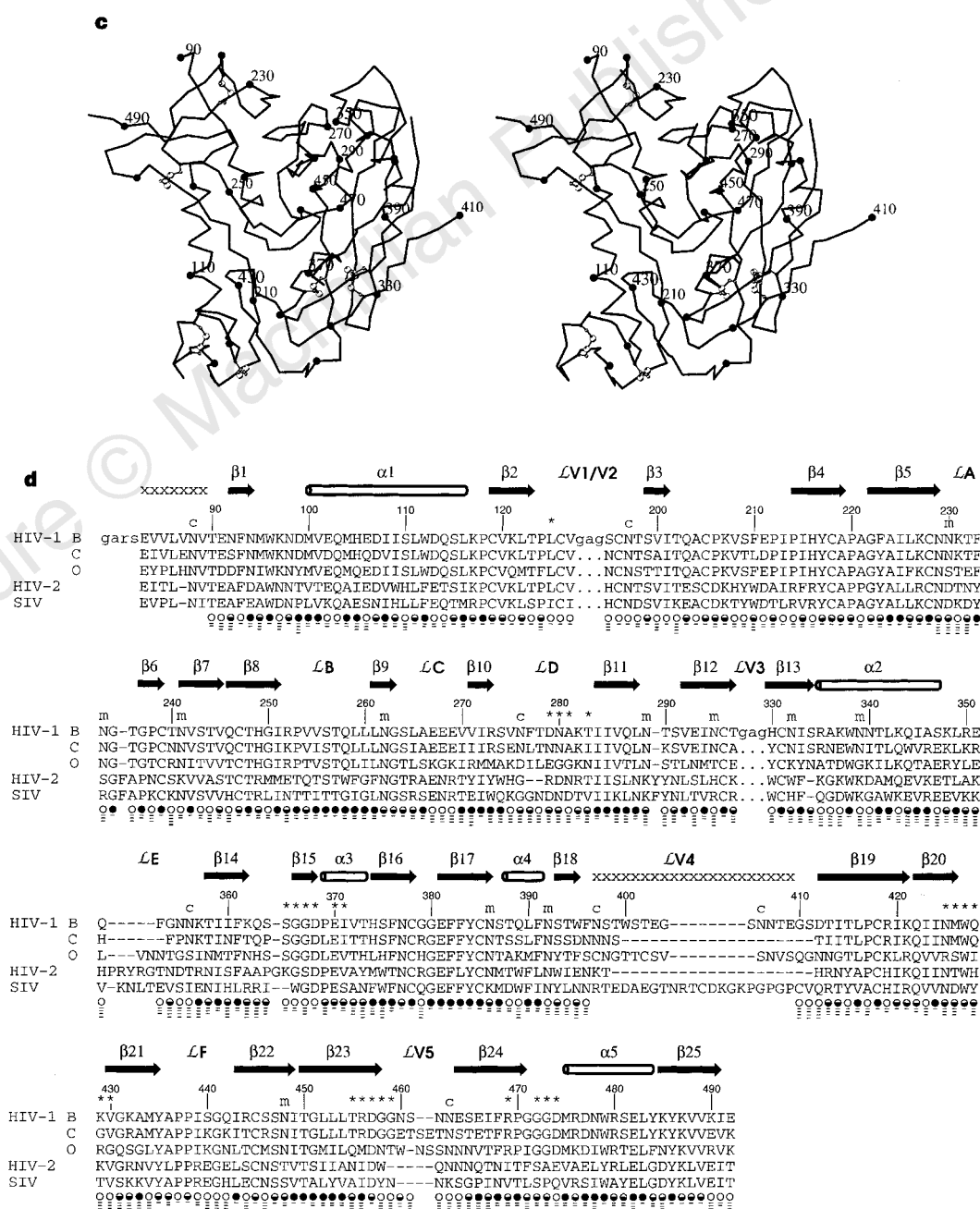
and HIV-2 comparisons, respectively. Variability correlates with the degree of solvent exposure of residues (Fig. 2d), in keeping with the conservation of hydrophobic cores. The seven disulphide bridges retained in core gp120 are absolutely conserved and mostly buried (Fig. 2c). Glycosylation sites are all surface-exposed and are conserved more than average (Fig. 2d). The previously identified HIV variable segments⁵ are all on loops connecting elements of secondary structure, and loops $\mathcal{L}\mathcal{D}$ and $\mathcal{L}\mathcal{E}$ are also especially variable. Indeed, $\mathcal{L}\mathcal{E}$ is more variable than V5 in light of current sequence data. These loops are also relatively mobile, as reflected in high B-factors or disorder, as in V4. Variable segments in the outer domain, including the exposed face of $\alpha 2$, appear to arise from neutral mutation rather than selective pressure, because they are on non-immunogenic surfaces²⁴, presumably masked by glycosylation.

CD4-gp120 interaction

CD4 is bound into a depression formed at the interface of the outer

domain with the inner domain and the bridging sheet of gp120 (Fig. 3a). This interaction buries a total of 742 Å² from CD4 and 802 Å² from gp120. The surface areas that are actually in contact are considerably smaller (Fig. 3d) because an unusual mismatch in surface topography creates large cavities that are occluded in the interface, as described below. There is, however, a general complementarity in electrostatic potential at the surfaces of contact, although the match is imprecise in this respect as well. The focus of CD4 positivity is displaced from the centre of greatest negativity on gp120 (Fig. 3c). The binding site is devoid of carbohydrate (Fig. 3g). The structure of CD4 in this complex differs only locally from that in free D1D2 structures and at only a few places: residues 17–20 at the poorly ordered CDR1-like loop and residues 41, 42, 47, 49 and 60, which are at or near the contact site and have low B factors in the gp120-bound state.

Direct interatomic contacts are made between 22 CD4 residues and 26 gp120 amino-acid residues. These include 219 van der Waals



contacts and 12 hydrogen bonds. Residues in contact are concentrated in the span from 25 to 64 of CD4, but they are distributed over six segments of gp120 (Figs 2d and 3i): one residue from the V1/V2 stem, loop $\mathcal{L}D$, the $\beta 15$ – $\alpha 3$ excursion, the $\beta 20$ – $\beta 21$ hairpin, strand $\beta 23$ and the $\beta 24$ – $\alpha 5$ connection. These interactions are compatible with previous analyses of mutational data on both CD4

(refs 11, 12, 29) and gp120 (refs 3, 13, 14). Other groups are also involved, including some at gp120 sites that have not been tested, but residues identified as critical for binding do indeed interact with one another (Fig. 3e). Most important, Phe 43 and Arg 59 of CD4 make multiple contacts centred on residues Asp 368, Glu 370 and Trp 427 of gp120, which are all conserved among primate

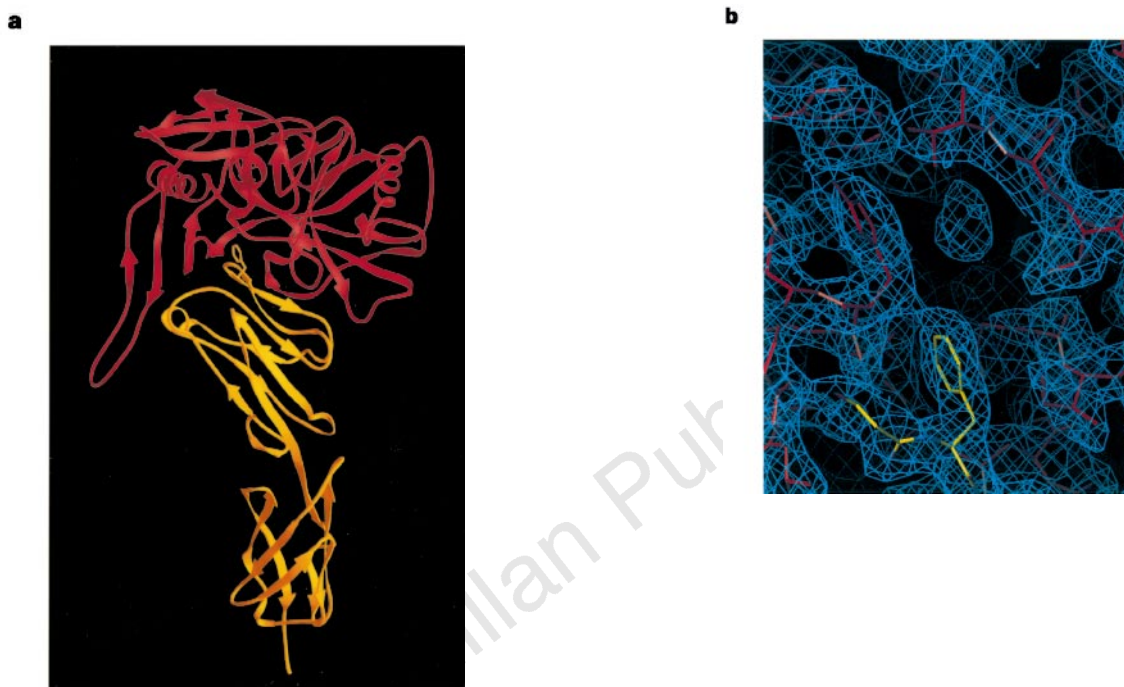


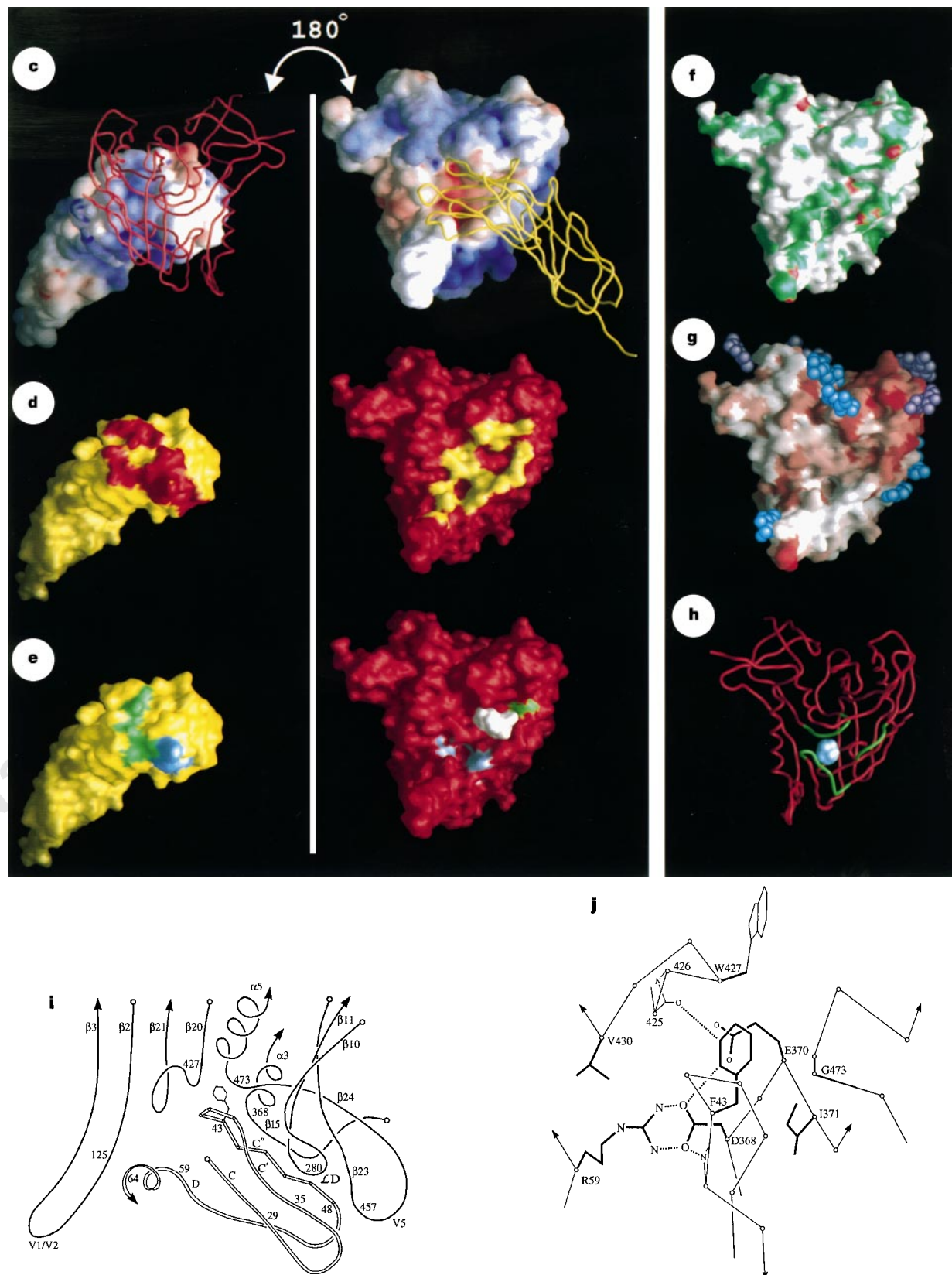
Figure 3 CD4–gp120 interactions. **a**, Ribbon diagram of gp120 (red) binding to CD4 (yellow). Residue Phe 43 of CD4 is also depicted reaching into the heart of gp120. From this orientation, the recessed nature of the gp120 binding pocket is evident. **b**, Electron density in the Phe 43 cavity. The $2F_o - F_c$ electron density map at $2.5 \text{ \AA}$, 1.1σ contour, is shown in blue. The gp120 model is depicted in red with the CD4 model in yellow (carbon atoms), blue (nitrogen atoms), and red (oxygen atoms). The orientation is the same as in **a**. The foreground has been clipped for clarity, removing the overlying $\beta 24$ – $\alpha 5$ connection. In the upper middle region lies the central unidentified density. At the bottom, Phe 43 of CD4 reaches up to contact the cavity. Moving clockwise round the cavity, the gp120 residues are Trp 427 (with its indole ring partially clipped by foreground slabbing), Trp 112, Val 255, Thr 257, Glu 370 (packing under the Phe 43 ring), Ile 371 and Asp 368 (partially clipped in the bottom right corner). Hydrophobic residues lining the back of the cavity can be partially seen around the central unidentified density. **c**, Electrostatic surfaces of CD4 and gp120. The electrostatic potential is shown at the solvent-accessible surface, which is coloured according to the local electrostatic potential, ranging from dark blue (most positive) to deep red (most negative). The slight puffiness of the surface arises from the enlarged solvent-accessible surface relative to the standard molecular surface. On the right, the gp120 surface is shown in an orientation similar to that in Fig. 2a, c, but rotated $\sim 20^\circ$ around a vertical axis to depict the recessed binding pocket more clearly. A thin yellow $\text{C}\alpha$ worm of CD4 is shown to aid in orientation. On the left, the CD4 surface is shown, rotated relative to the gp120 panel by an exact 180° rotation about the vertical axis shown. A thin red $\text{C}\alpha$ worm of gp120 is shown. **d**, CD4–gp120 contact surface. On the right, the gp120 surface is shown in red, with the surface within $3.5 \text{ \AA}$ of CD4 (surface-to-atom centre distance) in yellow. This effectively creates an imprint of CD4 on the gp120 surface. On the left (180° rotation), the corresponding CD4 surface is shown in yellow, with the gp120 imprint in red. **e**, CD4–gp120 mutational hot-spots. On the right, the surface of gp120 is shown in red, with the surface highlighted of gp120 residues that have been shown by substitution to affect CD4 binding: cyan, substantial effect—residues 368, 370 and 427; green, moderate effect—residue 457. Also depicted (white) is the surface of the large water-filled cavity at the CD4–gp120 interface. On the left (180° rotation), residues important for gp120 binding are shown on a

yellow CD4 surface: cyan, substantial effect—residues 43 and 59; green, moderate effect—residues 29, 35, 44, 46, 47 (ref. 29). **f**, Side-chain/main-chain contribution to the gp120 surface. The surface of gp120 contributed by main-chain atoms (including $\text{C}\beta$) is green, that contributed by side-chain atoms is white, and that contributed by the $\text{C}\alpha$ of glycine is brick-red. This orientation is the same as the right panel of **c–e** and in **g**, and allows for direct comparison of the CD4–gp120 contact surface. A striking surface concentration of main-chain atoms is seen in the regions corresponding to the CD4 imprint. **g**, Sequence variability mapped to the gp120 surface. The sequence variability among primate immunodeficiency viruses (Fig. 2d) is mapped onto the gp120 surface. A sliding scale of white (conserved) to brick-red (highly variable) is shown. Carbohydrate residues are also shown: blue, *N*-acetylglucosamine and fucose residues in the structure; purple, Asn-proximal *N*-acetylglucosamines modelled at residues 88, 230, 241, 356, 397, 406, 462. Much of the carbohydrate (22 residues) is hidden on the back side of the outer domain. **h**, Phe 43 cavity. The surface of the Phe 43 cavity is in blue, buried in the heart of gp120. A $\text{C}\alpha$ worm representation of gp120 (red) shows in green the three stretches that are incorrect by secondary-structure prediction: the $\mathcal{L}B$ loop, bending around the top of the cavity, parts of $\beta 20$ – $\beta 21$ just below the cavity, and strand $\beta 15$, slightly right of the cavity. The orientation shown here is the same as for the gp120 surfaces in **c–g**. **i**, The CD4–gp120 interface. This schematic representation of the entire interface shows six discrete segments of gp120 (solid black lines) interacting with CD4 (double lines). For orientation, secondary structural elements are labelled, as are representative contact residues from each segment of gp120. Arrows indicate main-chain direction. The side chain of Phe 43 is also shown. The orientation shown is similar to that in **a** and **b**. **j**, gp120 contacts around Phe 43 and Arg 59 of CD4. Residues on gp120 involved in direct contact with Phe 43 or Arg 59 are shown. Electrostatic interactions are depicted as dashed lines. Hydrophobic interactions are found between Phe 43 (CD4) and Trp 427, Glu 370, Gly 473 and Ile 371 (all from gp120) and between Arg 59 (CD4) and Val 430 (gp120). The orientation is similar to that in **a**, **b** and **i**, but has been rotated for clarity. Side chains of Phe 43 and Arg 59, as well as those portions of gp120 sidechains that interact with these crucial CD4 residues, are drawn as bold lines. Panel **a** was drawn with RIBBONS⁴⁹, **b** with program O⁴⁷, and **b–g** with GRASP⁵⁰.

immunodeficiency viruses. In fact, 63% of all interatomic contacts come from one span (40–48) in C'C'' of CD4, and Phe 43 alone accounts for 23% of the total. Similarly, with respect to gp120, the spans of 365–371 and 425–430 contribute 57% of the total. Of the three CD4 lysine residues implicated in binding (residues 29, 35 and 46), only Lys 29 makes a direct ionic hydrogen bond, and

although Asp 457 of gp120 is near to these electropositive groups (Fig. 3e, i), it does not make hydrogen bonds.

Several gp120 residues that are covered by CD4 are variable in sequence. This variation is accommodated in part by the large interfacial cavity (Fig. 3e). The gp120 residues in contact with this water-filled cavity are especially variable (Fig. 3g). Moreover, half of



the gp120 residues that make contacts with CD4 do so only through main-chain atoms (including C β) of gp120, and 60% of CD4 contacts are made by gp120 main-chain atoms (Fig. 3f). Included among these are 5 of the 12 hydrogen bonds in the interface. One such contributing element is an antiparallel β -sheet alignment of CD4 strand C' with gp120 strand β 15 (Fig. 3a, i).

Atomic details of the interaction are particularly intricate and unusual for the contacts made between gp120 and the mutationally critical CD4 residues Phe 43 and Arg 59 (Fig. 3j). Arg 59 interacts with Asp 368 and Val 430. The carboxylate group of Asp 368 makes double hydrogen-bonds with the guanidinium N η atoms of Arg 59, but it also hydrogen-bonds back to the backbone NH group of residue 44 and it appears to be optimally positioned to receive a CH...O hydrogen bond (3.20 Å) from the Phe 43 ring. Phe 43 interacts with residues Glu 370, Ile 371, Asn 425, Met 426, Trp 427 and Gly 473, as well as Asp 368, but only the contacts with Ile 371 have a conventional hydrophobic character. Those to 425–427 and 473, including Trp 427, are only to backbone atoms. A surprisingly large fraction of the Phe 43 contacts (28%) are to polar atoms. The phenyl group is stacked on the carboxylate group of Glu 370, and there are contacts with the carbonyl oxygen atoms of residues 425, 426 and 473 and the NH group of Trp 427. One of these, from the phenyl ring to O 425, is a second candidate CH...O hydrogen bond (3.10 Å). Asp 368 and Glu 370 have their carboxylate groups close together (3.54 Å) and they are, of course, buried in the complex. Even for gp120 excised from the complex, their fractional surface accessibilities are only 44 and 14%, respectively. Glu 370 may therefore be protonated. Perhaps the most extraordinary aspect of this site is the large cavity beyond C ζ of Phe 43 (Fig. 3b, h, and see below).

Interfacial cavities

Analysis of the solvent-accessible surface of the ternary complex reveals a number of topologically interior surfaces or cavities. Two of these, both at the gp120–CD4 interface, are unusually large. The larger (279 Å³) is formed at the interface between the slightly concave middle of the CC'C' portion of the CD4 sheet and a groove on gp120, where β 23 and β 24 are indented relative to β 15 and the LD loop (Fig. 3e). The second is from a pocket in the gp120 surface that is plugged by Phe 43 from CD4 (152 Å³). This pocket is itself at the interface between the inner and outer domains of gp120 (Fig. 3h). Several other smaller cavities are also wedged at the interface between the two gp120 domains.

The larger cavity is lined mostly by hydrophilic residues, half derived from gp120 and half from CD4. It is not deeply buried; although it is formally a cavity in the crystal structure, minor changes in side-chain orientation could make it solvent-accessible. The observed electron density and predicted hydrogen bonding are consistent with at least eight water molecules in the cavity. Residues from gp120 that actually line the cavity (including Ala 281, Ser 364, Ser 365, Thr 455, Arg 469) show sequence variability, whereas surrounding this variable patch are conserved residues, the substitution of which affect CD4 binding. These include the critical contact residues Asp 368, Glu 370 and Trp 427, which flank one end of the cavity, and Asp 457 at the other end (Fig. 3e). Similarly, CD4 residues that line the cavity (such as Gln 40 and Lys 35) can be mutated with only moderate effect on gp120 binding, whereas Arg 59 suffers less loss of solvent-accessible surface upon gp120 binding but is highly sensitive to mutation. This cavity thus serves as a water buffer between gp120 and CD4 (Fig. 3e). The tolerance for variation in the gp120 surface associated with this cavity produces a variational island (Fig. 3g), or 'anti-hotspot', which is centrally located between regions required for CD4 binding, and may help the virus escape from antibodies directed against the CD4-binding site.

The 'Phe 43' cavity (Fig. 3b, h) is very different in character from the larger binding–interface cavity. It is roughly spherical, with a

diameter of ~ 8 Å (atom centre to atom centre) across the centre of the cavity. It is positioned just beyond Phe 43 of CD4, at the intersection of the inner domain, the outer domain and the bridging sheet. It is deeply buried, extending into the hydrophobic interior of gp120. The phenyl ring of Phe 43 is the only non-gp120 residue contacting this cavity, forming a lid that covers the bottom of the cavity (Fig. 3b). Other routes of solvent access are possible: past Met 426 under the bridging sheet, or directly through the heart of gp120, at the interface between the inner and outer domains. Ordered water molecules demarking possible paths of solvent access are found along both routes. Nonetheless, in the cavity itself only a few water molecules are seen. The centre of the cavity is dominated by a large piece of spherical density, which is over 4 Å from any protein atom (Fig. 3b). The size, shape and predicted hydrogen bonding of this density is inconsistent with those expected for water, isopropanol, ethylene glycol, or any of the other major crystallization components. We have not identified the source of this density.

Residues that line the Phe 43 cavity (Trp 112, Val 255, Thr 257, Glu 370, Phe 382, Tyr 384, Trp 427 and Met 475; only main chains of 256 and 375–377) are primarily hydrophobic. They are also highly conserved, as much so as the buried gp120 hydrophobic core. Despite a lack of steric hindrance, almost no substitutions to larger residues are found. Given the frequency of gp120 sequence divergence, such conservation strongly implies functional significance. Indeed, although residues that line this cavity provide little direct contact to CD4, they do affect the gp120–CD4 interaction. Thus, mutations at Thr 257 (no contacts) and Trp 427 (only main-chain contacts) can substantially reduce binding. Changes in cavity-lining residues also affect the binding of antibodies directed against the CD4-binding site²⁴. In addition, many of the residues that line the cavity interact with elements of the chemokine-receptor-binding region (see below). It may be that the Phe 43 cavity and the other interdomain cavities form as a consequence of a CD4-induced conformational change (see below).

Despite the unusual cavity-laden interface between gp120 and CD4, we believe that this structure reflects the true character of the interaction. Core and full-length gp120 bind CD4 with similar affinity²⁶, and residues identified as critical by mutational analysis on both components are indeed at the focus of contact in the structure. In any case, the missing gp120 loops and termini could not conceivably have a role in filling these cavities.

Antibody interface

The 17b antibody is a broadly neutralizing human monoclonal isolated from the blood of an HIV-infected individual. It binds to a CD4-induced (CD4i) gp120 epitope that overlaps the chemokine-receptor-binding site²⁰.

Relative to other antibody–antigen pairs the interface between Fab 17b and core gp120 in the ternary complex is small (Fig. 4a–c). The solvent-accessible area excluded upon binding is only 455 Å² from gp120 and 445 Å² from 17b, which is largely from the heavy chain (371 Å²). The long (15 residue) complementarity-determining region 3 (CDR3) of the heavy chain dominates, but the heavy-chain CDR2 and the light-chain CDR3 also contribute. Overall, the 17b contact surface is very acidic (3 Asp, 3 Glu, no Arg or Lys), although hydrophobic contacts (notably a *cis* proline and tryptophan from the light chain) predominate at the centre.

On gp120, the 17b epitope lies across the base of the four-stranded bridging sheet (Fig. 4c, e). All four strands make substantial contact with 17b, suggesting that the integrity of the bridging sheet is necessary for 17b binding. The gp120 surface that contacts 17b consists of a hydrophobic centre surrounded by a highly basic periphery (3 Lys, 1 Arg, and no Asp or Glu) (Fig. 4d). Although this basic gp120 surface complements the acidic 17b surface, only one salt bridge is observed (between Arg 419 of gp120 and Glu 106 of the 17b heavy chain). The rest of the specific

contacts occur between hydrophobic and polar residues. Thus, the interaction between 17b and gp120 involves a hydrophobic central region flanked on the periphery by charged regions, predominately acidic on 17b and basic on gp120. There are no direct CD4–17b contacts and none of the gp120 residues contacts both 17b and CD4. Rather, CD4 binds on the opposite face of the bridging sheet, providing specific contacts that seem to stabilize its conformation (Fig. 3i, j) and may explain, in part, the CD4 induction of 17b binding.

The 17b epitope is well conserved among HIV-1 isolates. Of the 18 residues that show a loss in solvent-accessible surface upon contact with 17b, 12 residues (67%) are conserved among all HIV-1 viruses. By contrast, only 19 of the 37 gp120 residues (51%) that show loss of solvent-accessible surface upon CD4 binding are similarly conserved. CD4i epitopes tend to be masked from immune surveillance by the adjacent V2 and V3 loops²⁴. In the complex structure, a large gap is seen between gp120 and the tips of the light-chain CDR1 and CDR2 loops. Pointing directly at this gap is the base of the V3 loop. In intact gp120, the variable loops may

need to be bypassed for access to the conserved structures in the bridging sheet. The 17b epitope may be further protected from the immune system by a CD4-induced conformational change (see below).

Chemokine-receptor site

The site of interaction with the chemokine receptor CCR5 overlaps with the 17b epitope³⁰. Both are induced upon CD4 binding and both involve highly conserved residues. Mutational analysis showed that the basic and polar gp120 residues (Lys 121, Arg 119, Lys 21, Gln 422) that contact the 17b heavy chain also are important for CCR5 interaction³⁰. The hydrophobic and acidic surface of the 17b heavy chain may mimic the tyrosine-rich, acidic N-terminal region of CCR5, which is important for gp120 binding and HIV-1 entry^{31,32}. Geometrically, this site is directed at the cellular membrane when gp120 is engaged by CD4. Electrostatic interactions between the basic surface of the bridging sheet and the acidic chemokine receptor (and possibly the acid headgroups in the

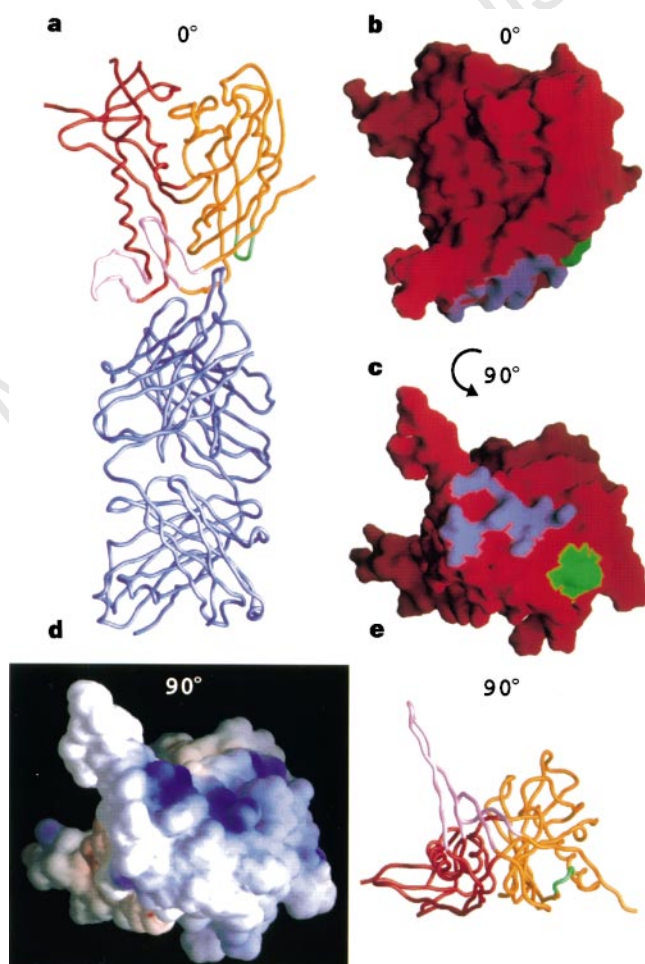


Figure 4 Neutralizing antibody 17b–gp120 interface. **a**, C α worm diagram of Fab 17b and gp120. The Fab 17b (blue) is shown binding to gp120, which has been coloured red (inner domain), pink (bridging sheet and V1/V2 stem), orange (outer domain) and green (V3 base). The orientation is the same as in Fig. 2a, c. **b**, Contact surface and V3 loop. The surface of gp120 is in red, with any surface within 3.5 Å of Fab 17b (surface-to-atom centre) coloured blue and the surface of the V3 base coloured green. The orientation is the same as in **a**. **c**, Contact surface and V3 loop. The same as **b** but rotated by 90° around a horizontal axis better to depict the 17b epitope. **d**, The electrostatic potential is shown at the

solvent-accessible surface, which is coloured according to the local electrostatic potential, ranging from dark blue (most positive) to red (negative). The electrostatic colouring is on the same scale as that in Fig. 3c. The surface that corresponds to the 17b epitope is the most electropositive region of the molecule. The V3 loop is truncated here, but sequence analysis shows it to be positively charged overall. **e**, C α -worm diagram of gp120. The gp120 is coloured according to the scheme in **a**. The orientation is the same as in **c** and **d**: that is, 90° from **a**. Figure was made with GRASP⁵⁰.

target membrane) could drive conformational changes related to virus entry.

Oligomer and gp41 interactions

Although monomeric in isolation, gp120 probably exists as a trimeric complex with gp41 on the virion surface. The large electroneutral surface on the inner domain (Fig. 3c) is the probable site of trimer packing based on its lack of glycosylation, its conservation in sequence, the location of CD4- and CCR5-binding sites, and the immune response to this region. These points are elaborated in the accompanying Letter and a model is presented²⁴.

Mutagenic and antibody-binding analyses indicate that the N and C termini of full-length gp120 are the most important regions for interaction with the gp41 glycoprotein^{33,34}. From these analyses, we expect that the gp41-interactive regions will extend away from core gp120 towards the viral membrane, and that the conserved, electroneutral surface is occluded in the oligomer/gp41 interface. A similar arrangement is seen in influenza haemagglutinin, where the extended N and C termini of HA₁ interact with the HA₂ transmembrane protein³⁵.

Conformational change in core gp120

There is abundant evidence to suggest that CD4 binding induces a conformational change in gp120, but much of it derives from intact gp120 with variable loops in place or from the oligomeric gp120–gp41 complex. The ternary complex structure provides clues to the nature of the conformational changes that occur within core gp120 itself. (Although 17b binding could contribute to the gp120 conformation observed in the crystal, the CD4 contacts are more extensive and multifaceted than those of 17b, indicating that CD4 binding may play the principal part in formation of the observed gp120 structure.)

Were the conformation of gp120 seen here preserved in the absence of CD4, the Phe 43 cavity (now a pocket) would present a perplexing structural dilemma. As discussed above, the cavity-lining residues have few structural restrictions, with ample room for larger substitutions into the cavity, yet these residues are highly conserved and inexplicably hydrophobic if exposed in a pocket. This pocket structure is in turn intimately connected to the bridging sheet, itself peculiar in absence of CD4. Thus, for example, the backbone amide of bridging-sheet residue 425 is hydrogen-bonded to Glu 370, a critical CD4 contact residue (Fig. 3j); Ile 424 makes extensive hydrophobic contacts with Phe 382, which lines the pocket from the outer domain; and Trp 427 packs perpendicular to Trp 112, which lines the pocket from the inner domain (Fig. 3b). Ne of Trp 427 is delicately poised for hydrogen bonding with the π -electrons of the indole ring of Trp 112. Structures such as these would necessarily be sensitive to orientational shifts between the

inner and outer domains.

The characteristics of 17b binding to core gp120 provide additional evidence for a CD4-induced conformational change. We do not observe detectable binding of Fab 17b to core gp120 unless CD4 is present, but then the ternary complex is stable in gel filtration. As there are no direct CD4–17b contacts in the structure, the effect of CD4 must be to stabilize the bridging-sheet minidomain to which 17b binds. This result is compatible with the binding properties of 17b and other CD4i antibodies to full-length gp120 (refs 18, 24), but it shows that the conformational change is not limited to an unmasking of the antibody epitope by CD4-induced displacement of the V2 loop, as was initially thought³⁶. The ability of the 17b antibody to bind full-length gp120 in the absence of CD4, albeit with lower affinity, implies that structural elements required for 17b binding can be accessed in the absence of CD4. If we assume that 17b binds in the same way to both full-length and core gp120, as shown by the concordance between the structural contacts (Fig. 4) and epitope-mapping data³⁰, then alternative conformations could be in a kinetically accessible equilibrium in native gp120.

A further indication that core gp120 may differ in the absence of CD4 comes from comparison with theory. When applied to the many known sequence variants of gp120, the evolutionary algorithm of PHD³⁷ gives secondary-structure predictions with estimated reliability above 90% for roughly 45% of the core gp120 sequence. Compared with our structure, it is accurate except at three places where it is markedly wrong (four consecutive residues with reliability index greater than 90%). All of these are at the Phe 43 cavity or in contacts with CD4: loop $\mathcal{L}B$, strand $\beta 15$, and the segment of $\beta 20$ into the turn to $\beta 21$ (Fig. 3h). Most significantly, the latter segment (residues 422–429) entering the bridging sheet is predicted to be helical. Indeed, residues 427–428 at the $\beta 20$ – $\beta 21$ turn do have helical character. We also note that CD4 binds efficiently to a gp120 derivative having both $\beta 2$ and $\beta 3$ truncated³⁸. As the bridging sheet is probably not stable in the absence of half of its strands, CD4 binding must be able to orientate strands $\beta 20$ and $\beta 21$ properly from a very different prior conformation.

The Phe 43 cavity is at the nexus of the CD4 interface, between the inner domain, the outer domain, and the bridging sheet. As such, Phe 43 itself seems to serve as a linchpin without which the structure might collapse. If so, to what state and, in reverse, how does CD4 binding lead to the state seen in this ternary complex? CD4–gp120 binding kinetics are complex³⁹, and microcalorimetry reveals unusually large ΔH and compensating $T\Delta S$ values for soluble CD4 binding to gp120 (M. L. Doyle, personal communication). These exceptional CD4-binding thermodynamics are suggestive of a large conformational change and are similar for both full-length and core gp120, supporting the relevance of our structural observations on core gp120. We imagine that CD4 sees gp120 as an uneven

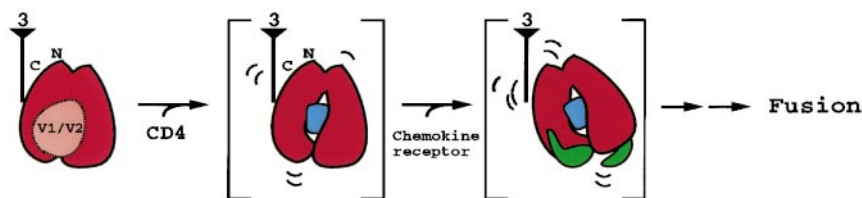


Figure 5 Diagram of gp120 initiation of fusion. A single monomer of core gp120 is depicted (red) in an orientation similar to that in Fig. 2a, c. The '3' symbolizes the 3-fold axis, from which gp41 interacts with the gp120 N and C termini to generate the functional oligomer. In the initial state of gp120 (on the surface of a virion), the V1/V2 loops (salmon) are shown partially occluding the CD4-binding site. Following CD4 binding (now at a target cell, though above the glycocalyx), a conformational change is depicted as an inner/outer domain shift, with the purple circle denoting

the formation of the Phe43 cavity. This conformational change strains the interactions at the N and C termini of gp120 with the rest of the oligomer, priming the CD4-bound gp120 core. In the next step (directly adjacent to the target membrane), the chemokine receptor binds to the bridging sheet and the V3 loop (in green; bottom left and right, respectively, of gp120), causing an orientational shift of core gp120 relative to the oligomer. This triggers further changes, which ultimately lead to the fusion of the viral and target membranes.

equilibrium of conformational states, makes initial contact by electrostatic interaction (Fig. 3c), stabilizes a nascent complex state, and inserts Phe 43 to weld the cavity into place.

Viral evasion of immune responses

Analysis of the antigenic structure of gp120 shows that most of the envelope protein surface is hidden from humoral immune responses by glycosylation and oligomeric occlusion²⁴. Most broadly neutralizing antibodies access only two surfaces: one that overlaps the CD4-binding site (shielded by the V1/V2 loop), and another that overlaps the chemokine-receptor-binding site (shielded by the V2 and V3 loops). Conformational changes in core gp120 provide additional mechanisms for evasion from immune surveillance. In the case of the CD4-binding surface, which contains a high proportion of main-chain atoms in the complex (Fig. 3f), the conformation without CD4 bound may expose underlying side-chain variability (Fig. 3g). Escape may also be provided by the recessed nature of the binding pocket (steric occlusion) (Fig. 3a) and by a topographical surface mismatch, which encloses a variational island or 'anti-hotspot' (described above; Fig. 3d). Mechanisms may be similar in the chemokine-receptor region: conformational change may hide the conserved epitope (unformed before CD4 binding); steric occlusion may take place between the CD4 anchored viral spike and the proximal target membrane; and an 'anti-hotspot' equivalent may camouflage chemokine-receptor-binding residues on the V3 loop in surrounding variability. Some of the means used to elude antibody-based defences may also help HIV avoid cellular immunity. Understanding the specific mechanisms of immune evasion may be prerequisite to the design of an effective vaccine.

Mechanistic implications for virus entry

During virus entry, HIV surface proteins fuse the viral membrane with the target cell membrane. gp120 is a crucial participant in the control and initiation of fusion. It functions in positioning: locating a cell capable of productive viral infection, anchoring the virus to the cell surface, and orienting the viral spike next to the target membrane. It also functions in timing: holding gp41 in a metastable conformation and triggering the coordinate release of the three N-terminal fusion peptides of the trimeric gp41. Although this is a complex, multiconformational process, the simplicity of the system, which comprises only two membranes, the viral oligomer and two host receptors, suggests that we may be able to understand the entire mechanism. Crystallography has now provided two snapshots: an intermediate state in which gp120 is bound to CD4, as described here, and a probably final, 'fusion-active' state of the gp41 ectodomain^{40,41}. Although atomic-level structures are lacking for other intermediates, the extensive biochemical data on membrane fusion mediated by HIV-1 envelope glycoproteins extend our understanding from these two states.

The entry process is initiated by binding of HIV-1 to the cellular receptor CD4 (Fig. 5, step 1). Although the extracellular portion of CD4 has some segmental flexibility, this binding roughly orients the viral spike. This orientation can be simulated by an alignment of the D1D2 portion of CD4 in the ternary complex with the previously solved structure of the four-domain, entire extracellular portion of CD4 (ref. 10). Such alignment orients the N and C termini of core gp120 towards the viral membrane, while the 17b epitope/chemokine-receptor-binding site on the gp120 surface faces the target cell membrane. Such an orientation is consistent with the proposed oligomeric structure and gp41-interactive surfaces (see above and ref. 24).

CD4 binding also induces conformational changes in gp120, which result in the creation of a metastable oligomer. Although some of the more flexible gp120 regions and gp41 are missing, the structure of the core gp120-CD4 complex presented here describes this state in atomic detail. CD4 binding results in movement of the V2 loop, which probably partially occludes the V3 loop and CD4i

epitopes^{18,36}. It also creates, or at least stabilizes, the bridging sheet on which the latter epitopes are located (described above). In addition, CD4 binding appears to change the V3 region, with enhanced proteolytic susceptibility at the tip of the loop and altered exposure of V3 epitopes¹⁹. The uncovered V3 and CD4i epitopes make up the chemokine-receptor binding site. Thus CD4 binding not only orients the gp120 surface implicated in chemokine receptor binding to face the target cell, it also forms and exposes the site itself. These changes may all happen in a single, concerted shift in the relative orientation of the inner and outer domains. This conformational shift may alter the orientation of the N and C termini, at the proximal end of the inner domain, perhaps partially destabilizing the oligomeric gp120/gp41 interface²¹. Such a shift would also alter the relative placement of the V1/V2 stem (in the CD4i site), which emanates from the inner domain, and the V3 loop, which emanates from the outer domain. Perhaps relevant to this, mutations that permit an adaptation of HIV-1 to CD4-independent entry using CXCR4 involve sequence changes in both the V1/V2 stem and the V3 loop⁴².

The next step in HIV-1 entry is the interaction of the gp120-CD4 complex with the chemokine receptor (Fig. 5, step 2). Although interactions between CD4 and chemokine receptor may occur, mutagenic analysis (H. Choe and J. Sodroski, unpublished observations) and the known examples of CD4-independent virus entry or chemokine-receptor binding suggest that direct gp120 contacts dominate interaction with the chemokine receptor. As most of the chemokine receptor is encased in the host membrane, binding will move the gp120 bridging sheet close to the target membrane. This movement must require CD4 flexibility because the initial HIV binding at the N-terminal D1 domains probably occurs above the glycocalyx. Reducing flexibility at the D2-D3 juncture or at the D4-membrane juncture of CD4 has been shown to block HIV-1 entry^{10,43}.

Chemokine-receptor binding is believed to trigger additional conformational changes in the HIV-1 envelope glycoprotein trimer that lead to exposure of the gp41 ectodomain. Presumably, a signal is transmitted from the cell-associated distal end of gp120 to elements of the inner domain that are likely to be involved in gp120-gp41 or gp120-gp120 association on the trimer. Although further interdomain shifts may occur in core gp120 after chemokine-receptor binding, the geometrically specific contacts that support the bridging sheet make it unlikely that another shift would occur without destabilizing this important component of the chemokine-receptor-binding site. As the high affinity of the interactions make it likely that both CD4 and chemokine receptor remain bound to gp120 during fusion, we expect additional conformational changes to occur between neighbouring gp120 protomers in the oligomeric complex. Perhaps the chemokine receptor triggers gp41 exposure by prising gp120 protomers away from the trimer axis, thus exerting a torque on the gp120-gp41 interface. It is interesting that several of the substitutions that affect chemokine-receptor binding in the context of monomeric gp120 appear to induce gp120 dissociation in an oligomeric context³⁰.

The structure of the gp120/CD4/17b antibody ternary complex described here reveals many molecular aspects of HIV-1 entry, including the atomic structure of gp120, the explicit interactions with CD4, and the conserved site of binding for the chemokine receptor. Still unknown are details of the apo state of core gp120, the oligomeric structure, the interaction with the chemokine receptor, the conformational changes that trigger reorganization of the gp41 ectodomain, and the structural basis for insertion of the fusion peptide of gp41 into the target membrane. Further understanding will require snapshots of other intermediates. The conformational complexity and intricate domain associations of gp120, like those of reverse transcriptase⁴⁴, the other large HIV translation product, may reflect genome restrictions at the protein level like those that lead to overlapping reading frames at the transcription level. Multiply protected infection machinery is contained in these condensed

intricacies. Its mechanism frustrates host defences; understanding them may inspire medical intervention. □

Methods

Protein production, crystallization and data collection. The two-domain CD4 (D1D2, residues 1–182) was produced in Chinese hamster ovarian cells⁸, the monoclonal antibody 17b in an Epstein–Barr virus immortalized B-cell clone isolated from an HIV-1-infected individual and fused with a murine B-cell fusion partner¹⁸, and the core gp120 from *Drosophila* Schneider 2 lines under the control of an inducible metallothionein promoter²⁰. The various biochemical manipulations (such as deglycosylation for gp120 and papain digestion to produce Fab 17b), protein purification and ternary complex crystallization are described elsewhere²⁵. The best crystals were small needles of cross-section only 30–40 μm . These were crosslinked by vapour-diffusion glutaraldehyde treatment (C. J. Lusty, personal communication), equilibrated with cryoprotectant containing stabilizer (10% ethylene glycol with 10.5% monomethyl-PEG5000, 10% isopropanol, 50 mM NaCl, 100 mM citrate/HEPES buffer, pH 6.3), transferred into immiscible oil (Paratone-N; Exxon), suspended in a small ethylene loop at the end of a mounting pin and flash-frozen in a cryostat nitrogen stream at 100 K.

Diffraction data were collected at beamline X4A, Brookhaven National Laboratory, using phosphorimaging plates and a Fuji BAS2000 scanner. To

avoid overlap problems from the relatively high mosaicity ($\sim 1.0^\circ$), oscillation data were collected using a rotation axis that was off-set at least 30° from the 197 Å *c* axis. Although crystals initially diffracted to Bragg spacings of greater than 2 Å, *b* axis mosaicity and substantial radiation damage, despite cryogenic cooling, reduced the overall limit to ~ 2.5 Å. Data processing and reduction were achieved using DENZO and SCALEPACK⁴⁵ (Table 1).

Structure determination and refinement. To locate the position of the Fab 17b in the ternary complex crystals, rotational searches with 52 different Fab models were made with the program MERLOT (P. M. Fitzgerald). The Fab fragments were aligned by superposition of their variable domains to allow comparison of rotational solutions. Even though four models showed more than 10% discrimination between the highest and second-highest solutions, no consistent rotational solution was found. Discrimination between correct and incorrect solutions was achieved by using confirmatory searches with the variable portion of the Fab. This was successful with only one model, molecule B of 1hil. Rigid-body refinement of the 1hil solution (XPLOR⁴⁶), allowing each immunoglobulin domain to move independently, produced a Patterson correlation of 24.9%. To locate the position of the two-domain CD4, each of the top 100 possible rotational solutions with each of three different CD4 models (1cdi, 1cdh, 3cd4), were searched for a distinctive translation solution (AMOR; J. Navaza). The translation searches used the rigid body refined Fab as a partial structure to help discriminate the correct solution. Two distinctive solutions were found: the 25th rotational solution of 3cd4 gave a translation correlation of 0.171 (versus 0.128 for the second-highest translation solution), and the 61st rotational solution of 1cdh gave 0.149 (versus 0.140). These two solutions were virtually identical. Rigid-body refinement in XPLOR⁴⁶ gave a Patterson correlation of 7.9% for CD4 alone and 32.4% for Fab and CD4. All molecular replacement and rigid-body refinements used 8–4 Å data.

To provide additional phasing, crystals were soaked in over 20 different heavy-atom solutions and screened for isomorphous replacement using the statistical chi-squared test in SCALEPACK⁴⁵. Derivatives were identified from two heavy-atom compounds: 10 mM K_3IrCl_6 (10 h equilibration in a solution containing the heavy atom and cryoprotectant stabilizer; 2.8 Å) and 5 mM K_2OsCl_6 (24 h soak; 3.5 Å). Isomorphism was found to be highest between these heavy-atom data sets and a native data set collected at pH 7.0 (cryoprotectant stabilizer buffered with 50 mM bis-Tris, pH 7.0). Heavy-atom sites were identified by difference Fourier analysis using the molecular replacement phases, and phasing parameters were refined with MLPHARE (in the CCP4 suite of crystallographic programs). The K_3IrCl_6 derivative was modelled as nine partially occupied sites: two sites of occupancy 0.158 and 0.142, and seven of less than 0.07. Although relatively isomorphous, poor data quality (R_{sym} of more than 20% past 3.0 Å) combined with relatively small isomorphous differences (R_{iso} of 12.0%) reduced the quality of phasing. In contrast, the K_2OsCl_6 derivative had an R_{iso} of 15.6%, but was only isomorphous to ~ 5 Å. It was modelled as four sites of occupancy 0.321, 0.207, 0.194 and 0.128, with the highest site at the same position as the second-highest site from K_3IrCl_6 .

The initial combination of model and isomorphous replacement phasing did not produce readily interpretable density for gp120. To monitor phase improvement, we devised an objective assay of density quality that used correlations in a region internal to domain 1 of CD4 between the experimental electron density and the calculated model density (CD4 as positioned by molecular replacement and rigid-body refinement). Refinement of heavy-atom positions improved this correlation, and provided a starting point for phase improvement, primarily using real-space modification techniques (Table 1). These techniques included automatic concatenation of unmodelled density (with the program PRISM; D. Agard), reciprocal-space averaging of the PRISM modelled density and real-space model subtraction (implemented using the XPLOR⁴⁶ shell language), application of real-space constraints such as solvent flattening, histogram matching and negative-density truncation (with the program DM (in the CCP4 suite of crystallographic programs)), and real-space combined addition of the various experimental density maps (with the program MAPMAN; T. A. Jones). The combined application of these techniques greatly improved electron-density maps (Table 1).

At this point, most of the $\text{C}\alpha$ backbone could be modelled (with program O⁴⁷), thereby defining the secondary structure. Computer-aided sequence alignment (slider routine in O) and secondary-structure prediction (PHD³⁷) helped to position the amino-acid sequence, leaving only regions around the N

Table 1 Structure solution

Data collection	Native	K ₃ IrCl ₆	K ₂ OsCl ₆
Bragg spacing limits (Å)	20–2.5	20–2.8	20–3.5
Total observations	113,966	76,739	25,821
Unique observations	37,724	28,599	11,982
R _{sym} (%)††	9.3 (24.7)	11.5 (20.2)	14.3 (18.2)
Data coverage (%)†	86.0 (62.8)	90.8 (82.9)	72.5 (62.5)
Molecular replacement	Fab	CD4	Fab+CD4
Model	1hil	3cd4/1cdh	1hil+1cdh
Scattering (%)‡	43	18	61
Rigid-body correlation§	0.249	0.079	0.325
Generation of experimental electron density			
Phasing procedure	Correlation coefficient		
Molecular replacement (MR)	–0.02		
Multiple isomorphous replacement (MIR)	0.34		
Phase combination			
MIR + MR	0.60		
+ density modification	0.66		
+ density modification + subtraction	0.69		
Density modelling (concatenation)			
MIR + MR	0.65		
+ density modification	0.68		
+ density modification + subtraction	0.71		
Combinatorial map addition	0.73		
Refinement statistics			
R-values (10–2.5 Å):			
Data cutoff (σ)	0	2	4
R _{crystal} (R _{free}) (%)	24.9 (32.8)	22.2 (30.7)	21.2 (29.2)
Data completeness (%)	85.8	77.3	66.4
Geometric parameters (r.m.s.):			
Bond length (Å)	0.007		
Bond angle (°)	1.59°		
B-factors:	Average	r.m.s. _{bond}	r.m.s. _{angle}
main chain	20.80	1.33	2.31
side chain	21.93	1.97	3.01
water molecules	22.31		

* $R_{\text{sym}} = \sum |I_{\text{obs}} - I_{\text{avg}}| / \sum I_{\text{avg}}$.

† Numbers in parentheses represent the statistics for the shell comprising the outer 10% (theoretical) of the data.

‡ The percentage of scattering of the initial search model.

§ Correlation obtained upon rigid-body refinement of the model against 8–4 Å data.

|| Correlation in the D1 region of CD4 between the experimental electron density and the calculated model density (from CD4 as positioned by molecular replacement) using 10–2.8 Å data. Correlations in this region (consisting of $\sim 6,000$ Å³) were used to generate a quantitative measure of the overall quality of the ternary complex experimental electron density. For the purposes of these calculations, the model used for phase combination omitted D1. A correlation of 0.6 is roughly the level of an interpretable protein electron-density map; a well refined structure would give a correlation of ~ 0.9 .

terminus (residues 79–100 and residues 215–245), the V1/V2 loop, and the V4 loop uncertain. Iterative rounds of building with O, simulated annealing and positional refinement with XPLOR⁴⁶, and addition of ordered solvent clarified the trace.

Structure analysis. Deviations of the CD4 structure in the complex from the free state were measured as described³⁹. Deviations were taken as significant when the root-mean-square (r.m.s.) residue deviation was greater than the overall value and also more than 0.5 Å greater than variation among the free structures. Interatomic contacts were defined as in ref. 48. Structural alignments were made by visual comparison of the SCOP database, and automatic searches were made with PrISM (A.-S. Yang and B. Honig).

Received 15 May; accepted 28 May 1998.

- Barre-Sinoussi, F. *et al.* Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immunodeficiency syndrome (AIDS). *Science* **220**, 868–871 (1983).
- Gallo, R. C. *et al.* Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science* **224**, 500–503 (1984).
- Kowalski, M. L. *et al.* Functional regions of the envelope glycoprotein of human immunodeficiency virus type 1. *Science* **237**, 1351–1355 (1987).
- Lu, M., Blackow, S. & Kim, P. A trimeric structural domain of the HIV-1 transmembrane glycoprotein. *Nature Struct. Biol.* **2**, 1075–1082 (1995).
- Starcich, B. R. *et al.* Identification and characterization of conserved and variable regions of the envelope gene HTLV-III/LAV, the retrovirus of AIDS. *Cell* **45**, 637–648 (1986).
- Leonard, C. K. *et al.* Assignment of intrachain disulfide bonds and characterization of potential glycosylation sites of the type 1 recombinant immunodeficiency virus envelope glycoprotein (gp120) expressed in Chinese hamster ovary cell. *J. Biol. Chem.* **265**, 10373–10382 (1990).
- Profy, A. T. *et al.* Epitopes recognized by the neutralizing antibodies of an HIV-1-infected individual. *J. Immunol.* **144**, 4641–4647 (1990).
- Ryu, S.-E. *et al.* Crystal structure of an HIV-binding recombinant fragment of human CD4. *Nature* **348**, 419–426 (1990).
- Wang, J. H. *et al.* Atomic structure of a fragment of human CD4 containing two immunoglobulin-like domains. *Nature* **348**, 411–418 (1990).
- Wu, H., Kwong, P. D. & Hendrickson, W. A. Dimeric association and segmental variability in the structure of human CD4. *Nature* **387**, 527–530 (1997).
- Moebius, U., Clayton, L., Abraham, S., Harrison, S. & Reinherz, E. The human immunodeficiency virus gp120 binding site of CD4: Delineation by quantitative equilibrium and kinetic binding studies of mutants in conjunction with a high-resolution CD4 atomic structure. *J. Exp. Med.* **176**, 507–517 (1992).
- Sweet, R. W., Truneh, A. & Hendrickson, W. A. CD4: its structure, role in immune function and AIDS pathogenesis, and potential as a pharmacological target. *Curr. Opin. Biotech.* **2**, 622–633 (1991).
- Olshchey, U. *et al.* Identification of individual HIV-1 gp120 amino acids important for CD4 receptor binding. *J. Virol.* **64** (1990).
- Cordonnier, A., Montagnier, L. & Emerman, M. Single amino-acid changes in HIV envelope affect viral tropism and receptor binding. *Nature* **340**, 571–574 (1989).
- Moore, J. P. Coreceptors: implications for HIV pathogenesis and therapy. *Science* **276**, 51–52 (1997).
- Feng, E., Broder, C. C., Kennedy, P. E. & Berger, E. A. HIV-1-entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science* **272**, 872–877 (1996).
- Speck, R. F. *et al.* Selective employment of chemokine receptors as human immunodeficiency virus type 1 coreceptors determined by individual amino acids in the envelope V3 loop. *J. Virol.* **71**, 7136–7139 (1997).
- Thali, M. *et al.* Characterization of conserved human immunodeficiency virus type 1 (HIV-1) gp120 neutralization epitopes exposed upon gp120-CD4 binding. *J. Virol.* **67**, 3978–3988 (1993).
- Sattentau, Q. J., Moore, J. P., Vignaux, F., Traincard, F. & Poignard, P. Conformational changes induced in the envelope glycoproteins of human and simian immunodeficiency virus by soluble receptor binding. *J. Virol.* **64**, 7383–7383 (1993).
- Wu, L. *et al.* CD4-induced interaction of primary HIV-1 gp120 glycoproteins with the chemokine receptor CCR5. *Nature* **384**, 179–183 (1996).
- Moore, J. P., McKeating, J. A., Weiss, R. A. & Sattentau, Q. J. Dissociation of gp120 from HIV-1 virions induced by soluble CD4. *Science* **250**, 1139–1142 (1990).
- Bullough, P. A., Hughson, F. M., Skehel, J. J. & Wiley, D. C. Structure of influenza haemagglutinin at the pH of membrane fusion. *Nature* **371**, 37–43 (1994).
- Fass, D. *et al.* Structure of a murine leukemia virus receptor-binding glycoprotein at 2.0 Å resolution. *Science* **277**, 1662–1666 (1997).
- Wyatt, R. *et al.* The antigenic structure of the HIV gp120 envelope glycoprotein. *Nature* **393**, 705–711 (1998).
- Kwong, P. D. *et al.* Quantitative probability analysis and variational crystallization of gp120, the exterior envelope glycoprotein of the human immunodeficiency virus type 1 (HIV-1). *J. Biol. Chem.* (submitted).
- Binley, J. M. *et al.* Analysis of the interaction of antibodies with a conserved, enzymatically deglycosylated core of the HIV-1 gp120 envelope glycoprotein. *AIDS Res. Hum. Retrovir.* **14**, 191–198 (1997).
- Leesong, M., Hederson, B. S., Gillig, J. R., Schwab, J. M. & Smith, J. L. Structure of a dehydratase-isomerase from the bacterial pathway for biosynthesis of unsaturated fatty acids: two catalytic activities in one active site. *Structure* **4**, 253–264 (1996).
- Cederger-Zeppezauer, E. S., Larsson, G., Nyman, P. O., Dauter, Z. & Wilson, K. S. Crystal structure of a dUTPase. *Nature* **355**, 740–743 (1992).
- Ryu, S.-E., Truneh, A., Sweet, R. W. & Hendrickson, W. A. Structures of an HIV and MHC binding fragment from human CD4 as refined in two crystal lattices. *Structure* **2**, 59–74 (1994).
- Rizzuto, C. *et al.* A conserved HIV gp120 glycoprotein structure involved in chemokine receptor binding. *Science* (in the press).
- Dragic, T. *et al.* Amino-terminal substitutions in the CCR5 coreceptor impair gp120 binding and human immunodeficiency type 1 entry. *J. Virol.* **72**, 279–285 (1998).
- Farzan, M. *et al.* A tyrosine-rich region in the N-terminus of CCR5 is important for human immunodeficiency virus type 1 entry and mediates an association between gp120 and CCR5. *J. Virol.* **72**, 1160–1164 (1998).
- Wyatt, R. *et al.* Analysis of the interaction of the human immunodeficiency virus type 1 gp120 envelope glycoprotein with the gp41 transmembrane glycoprotein. *J. Virol.* **71**, 9722–9731 (1997).
- Helseth, E., Olshevsky, U., Furman, C. & Sodroski, J. Human immunodeficiency virus type 1 gp120 envelope glycoprotein regions important for association with the gp41 transmembrane glycoprotein. *J. Virol.* **65**, 2119–2123 (1991).
- Wilson, I. A., Skehel, J. J. & Wiley, D. C. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. *Nature* **289**, 366–373 (1981).
- Wyatt, R. *et al.* Involvement of the V1/V2 variable loop structure in the exposure of human immunodeficiency virus type 1 gp120 epitopes induced by receptor binding. *J. Virol.* **69**, 5723–5733 (1995).
- Rost, B., Sander, C. & Schneider, R. PHD—an automated mail server for protein secondary structure prediction. *Comput. Appl. Biosci.* **10**, 53–60 (1994).
- Wyatt, R. *et al.* Functional and immunologic characterization of human immunodeficiency virus type 1 envelope glycoproteins containing deletions of the major variable regions. *J. Virol.* **67**, 4557–4565 (1993).
- Wu, H. *et al.* Kinetic and structural analysis of mutant CD4 receptors that are defective in HIV gp120 binding. *Proc. Natl. Acad. Sci. USA* **93**, 15030–15035 (1996).
- Chan, D. C., Fass, D., Berger, J. M. & Kim, P. S. Core structure of gp41 from the HIV envelope glycoprotein. *Cell* **89**, 263–273 (1997).
- Weissenhorn, W., Dessen, A., Harrison, S. C., Skehel, J. J. & Wiley, D. C. Atomic structure of the ectodomain from HIV-1 gp41. *Nature* **387**, 426–430 (1997).
- Dumonceaux, J. *et al.* Spontaneous mutations in the env gene of the human immunodeficiency virus type 1 NDK isolate are associated with a CD4-independent entry phenotype. *J. Virol.* **72**, 519–519 (1998).
- Moir, S., Perreault, J. & Poulin, L. Postbinding events mediated by human immunodeficiency virus type 1 are sensitive to modifications in the D4-transmembrane linked region of CD4. *J. Virol.* **70**, 8019–8028 (1996).
- Kohlstaadt, L. A., Wang, J., Friedman, J. M., Rice, P. A. & Steitz, T. A. Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science* **256**, 1783–1790 (1992).
- Otwiński, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Meth. Enzymol.* **276**, 307–326 (1997).
- Brünger, A. T. *XPLOR Version 3.1 manual* (Yale University, New Haven, 1993).
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Zhu, X. *et al.* Structural analysis of substrate binding by the molecular chaperone DnaK. *Science* **272**, 1606–1614 (1996).
- Carson, M. *Ribbons 2.0. J. Appl. Crystallogr.* **24**, 958–961 (1991).
- Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insight from the interfacial and thermodynamic properties of hydrocarbons. *Proteins Struct. Funct. Genet.* **11**, 281–296 (1991).

Acknowledgements. We thank M. Doyle for microcalorimetry results, C. Kokolis and the SSCC group at SmithKline Beecham for assistance with gp120 proteins, C. Lusty for suggestions on crosslinking, C. Ogata for beamline assistance, A.-S. Yang for automated structural alignment, and past and present members of W.A.H.'s group, especially D. Fremont for help with programs, E. Martinez-Hackert for help with figures, L. Shapiro for discussion, J. Williams for sequence alignments, and H. Wu for analysis of CD4 movement. Beamline X4A at the National Synchrotron Light Source, a Department of Energy facility, is supported by the Howard Hughes Medical Institute. This work was made possible by gifts from the late William McCarty-Cooper, from the G. Harold and Leila Y. Mathers Foundation, from the Friends 10, and from Douglas and Judi Krupp. Support was provided by the Aaron Diamond Foundation, the American Foundation for AIDS Research, the Howard Hughes Medical Institute, and the NIH.

Correspondence and requests for materials should be addressed to W.A.H. (e-mail: wayne@convex.hhmi.columbia.edu). Coordinates have been deposited in the Brookhaven Protein Data Bank (accession code 1gc1) and may be obtained from the authors.

A large-scale, interstellar Faraday-rotation feature of unknown origin

A. D. Gray*, T. L. Landecker*, P. E. Dewdney* & A. R. Taylor†

* National Research Council Canada, Herzberg Institute of Astrophysics, Dominion Radio Astrophysical Observatory, Box 248, Penticton, British Columbia, V2A 6K3, Canada

† Department of Physics and Astronomy, University of Calgary, 2500 University Drive NW, Calgary, Alberta, T2N 1N4, Canada

The disk of the Milky Way contains free electrons and magnetic fields which contribute significantly to the energetics of the interstellar medium¹. The concentrations of electrons and magnetic fields are too low to be detected by direct methods, but may be investigated using Faraday rotation, a wavelength-dependent shift in linear polarization angle induced by a magneto-ionic medium². Structures in polarization angle arising from Faraday rotation have been detected recently at long radio wavelengths³. These structures are disorganized and filamentary, probably arising from interstellar gas in the vicinity of the Sun. Here we report a more distant, highly ordered Faraday-rotation structure of elliptical shape, with its long axis parallel to the plane of the Galaxy. The feature appears to be located in an inter-arm region of the Milky Way, between the spiral arm containing the Sun and the next outer (Perseus) spiral arm. Within the elliptical region, small-scale structure which characterizes the turbulence seen in adjacent regions of the interstellar medium is absent. The origin of this magneto-ionic feature is uncertain, but it must arise from an organization of the magnetic-field and electron-density distributions on a scale of the order of 50 parsecs (165 light years).

The data presented here form part of a 1,420 MHz (wavelength $\lambda = 21$ cm) radioastronomy investigation of the W3/W4/W5 star-formation complex⁴, which is located ~ 2.2 kpc away in the Perseus arm of the Milky Way. The observations were made with the Synthesis Telescope at the Dominion Radio Astrophysical Observatory, which produces interferometric images of 2° -diameter fields, sensitive to spatial components from $1'$ to $\sim 1^\circ$ with an r.m.s. noise of 0.3 mJy per beam. The polarimeter is sensitive to orthogonal circular polarizations; instrumental polarization is calibrated by observations of unresolved radio sources with known polarization (3C147, 3C295 and 3C286)⁵. The residuals after calibration are 0.3% of Stokes I (total intensity; see ref. 2).

Our observations have revealed a highly organized structure in linear polarization (Figs 1 and 2) over an elliptical region approximately $2^\circ \times 1^\circ$ in extent. Over this region the polarization angle decreases monotonically and symmetrically from a central extremum to the edge, with only small changes in polarized intensity. This indicates that the observed phenomenon arises from the Faraday rotation of a polarized radiation field by a foreground magneto-ionic medium (a Faraday screen). The background radiation here must be the highly polarized, diffuse Galactic synchrotron emission⁶. This emission is generally too smoothly distributed to be seen by interferometers, but the Faraday screen has here produced structure in polarization angle on scales to which the interferometer is sensitive.

Because of the unusual properties of this feature, we have investigated the possibility that arises from an instrumental effect. Along this line-of-sight, the Stokes I emission is dominated by the H II region W5 (contours in Fig. 2). There is no morphological resemblance between the smooth elliptical Faraday-rotation feature (FRF) and the irregular, filamentary structure of W5, as would be expected if the former were a polarized instrumental response.

Moreover, the strength of the polarized intensity is as much as 5% of I , well above the known instrumental levels of 0.3%. The FRF is also seen in both raw and processed images with identical structure at the same celestial position in images made of adjacent areas, whereas spurious effects due to off-field sources would vary in position or amplitude relative to celestial objects. This demonstrates conclusively that the feature has an astrophysical origin.

The change in polarization angle from the centre to the observable edge of the FRF is $\Delta\theta = 280^\circ$, requiring a change in rotation measure of $\Delta R_M = 110 \text{ rad m}^{-2}$. This is much smaller than could occur through the W5 region, for which we estimate $\Delta R_M \approx 2 \times 10^3 \text{ rad m}^{-2}$, assuming a typical value of magnetic field of $3 \mu\text{G}$ along with the known properties of W5 (ref. 7). Such high rotation measures have a strong depolarizing effect, due to spatial and spectral averaging of non-parallel polarization vectors⁸. This effect is evident in Figs 1 and 2, where the boundary of the H II region coincides with the disappearance of the surrounding polarization structure. Therefore the FRF cannot lie within or behind the W5 region. Furthermore, W5 has a sharp boundary⁹, but the FRF extends beyond it, also suggesting that the two are spatially distinct.

Apart from the requirement that the FRF lies no farther away than W5, there are no unambiguous indicators for its distance, nor of its dimension along the line of sight. For the following discussion we will use f to denote the distance to the FRF as a fraction of the distance to W5 (that is, $0 < f < 1$, where the total distance is 2.2 kpc), and assume that the phenomenon causing the FRF takes place in a volume whose depth is approximately the observed width ($80f \text{ pc}$).

The magnitude of the rotation measure change within the small volume of the FRF is large compared to that seen in an adjacent line of sight through the Galaxy. The radio galaxy 3C69 lies $\sim 1^\circ$ from the edge of the FRF and is known¹⁰ to have a rotation measure of $R_M = -179 \text{ rad m}^{-2}$. This rotation measure is typical of sources in this area of sky, and there are no magnetic field reversals along the path of propagation¹¹, and thus no changes in the sign of rotation measure. The mean rotation measure per unit distance through the FRF is thus $\sim 40f^{-1}$ times larger than the mean interstellar value in this direction.

The change in rotation measure across the FRF could be produced by variations in line-of-sight magnetic field strength, electron density, or a combination of both. In the first limiting case, assuming that the region has a constant electron density of typical interstellar value¹² $n_e \approx 0.1 \text{ cm}^{-3}$, the observations can be explained by a funnel-shaped magnetic field with line-of-sight strength varying by $\Delta B = 16f^{-1} \mu\text{G}$. As the average fluctuation in the line-of-sight component of the Galactic magnetic field¹³ is only $3 \mu\text{G}$, it is unlikely that the FRF arises from changes in the magnetic field alone; although evidence exists that in at least some spiral galaxies there are enhanced magnetic fields in the inter-arm regions^{14,15}, there is no evidence that this is the case for the Milky Way^{16,17}. We note that a symmetric change in the magnetic field emanating from a central object could not produce the observed effect, as the integrated line-of-sight component—and hence the rotation measure—along any path is constant in such a model.

For the second limiting case we assume an anomalous electron density inside an oblate spheroid, threaded by a constant magnetic field with a line-of-sight component $B_{||} \approx 3 \mu\text{G}$. The observed change in rotation measure implies a density difference inside the spheroid of $\Delta n_e = 0.5f^{-1} \text{ cm}^{-3}$. This could be either an excess or deficiency of electrons relative to the surrounding medium, depending on the magnetic field direction. However, the required change is at least 5 times the typical interstellar density. This can only be achieved via a deficiency if the external ambient density is high, and can only be sustained in pressure equilibrium if the space is filled with neutral gas. A pressure-equalized cloud of neutral hydrogen at a temperature of 100 K would contain a column density of $\sim 10^{22} \text{ cm}^{-2}$,

an order of magnitude higher than typical H I clouds, and should be easily detectable in 21-cm H I-line observations. If molecular hydrogen fills the void it would be even more dense, and detectable in CO-line emission. But both H I-line and CO-line observations fail to detect anything that is not associated with W5 (refs 4, 18). Therefore, we consider that an excess of electrons within the FRF is the most likely explanation.

An excess density of electrons is consistent with the suggested

origin of Faraday rotation fluctuations in the interstellar medium on scales of a few arcminutes or less, which are believed to be due to small (~ 1 pc) H II regions whose surface brightnesses are too low to detect directly¹⁹. However, here we are dealing with a structure perhaps two orders of magnitude larger in size. Integrating Δn_e over the volume of the FRF implies a total ionized hydrogen mass of $M_{\text{HII}} = 2 \times 10^3 f^2 M_{\odot}$ (where $M_{\odot}$ is the solar mass), and an excess emission measure of $\Delta E_M = 20 f^{-1} \text{ cm}^{-6} \text{ pc}$. For large values of f ,

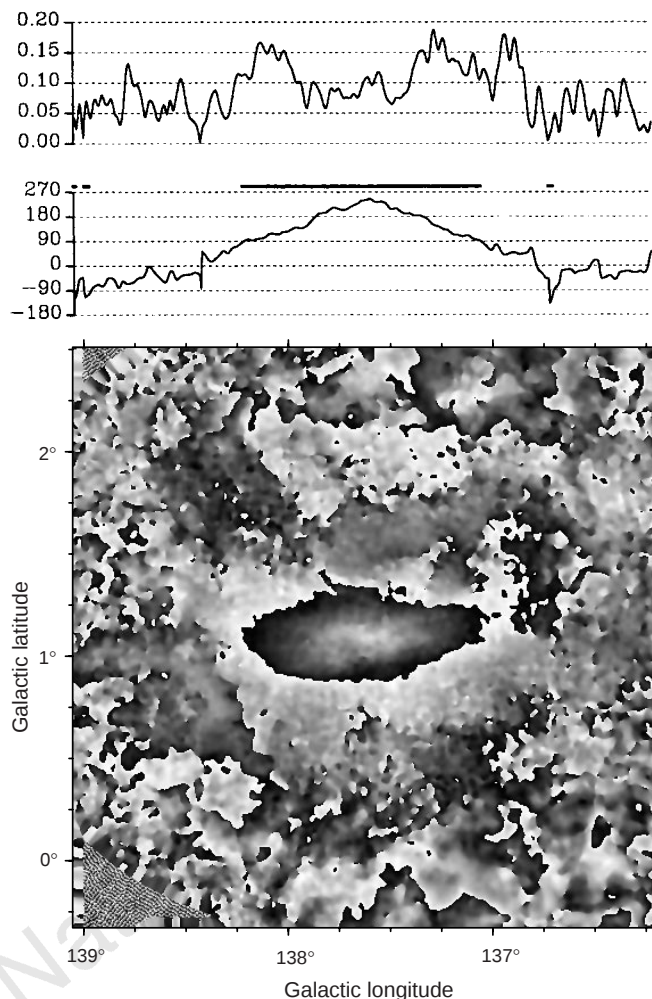


Figure 1 Grey-scale image of polarization angle made from DRAO data smoothed to $2'$. (Polarization angle $\theta_p = \frac{1}{2} \arctan(U/Q)$; Fig. 2 shows the corresponding Stokes Q and U images.) The grey-scale runs from $\theta_p = -90^\circ$ (black) to $\theta_p = +90^\circ$ (white). This image, presented in Galactic coordinates, is centred at longitude $l = 137.6^\circ$, latitude $b = +1.1^\circ$ (right ascension 02 h 54 min 50 s, declination $+60^\circ 24'$ (J2000)), and has been extracted from the edge of a larger montage of ten fields, which was formed by the noise-weighted sum of the overlapping images. The two line-plots above the image are slices at $b = +1.1^\circ$ through polarized intensity ($P = \sqrt{Q^2 + U^2}$; upper plot, in K) and polarization angle (lower, in degrees). The latter plot has had the angles 'unwrapped' to remove discontinuities arising from the $\pm 90^\circ$ ambiguity in θ_p . The main effect here was to remove the prominent angle wraps in the region of the FRF itself; the portions of data to which $\pm 180^\circ$ was added are indicated by the heavy horizontal line segments above the plot. The remaining discontinuities near $l = 136.7^\circ$ and $l = 138.4^\circ$ arise because the polarized intensity is very low at those points, and the uncertainty in polarization angle increases as polarized intensity decreases. In the region of the FRF the uncertainties are typically $\sim 20^\circ$. We note that the interferometer is sensitive only to changes in intensity, not intensity itself, and the variations in the upper plot represent a small fraction of the total polarized intensity (~ 1 K; see text).

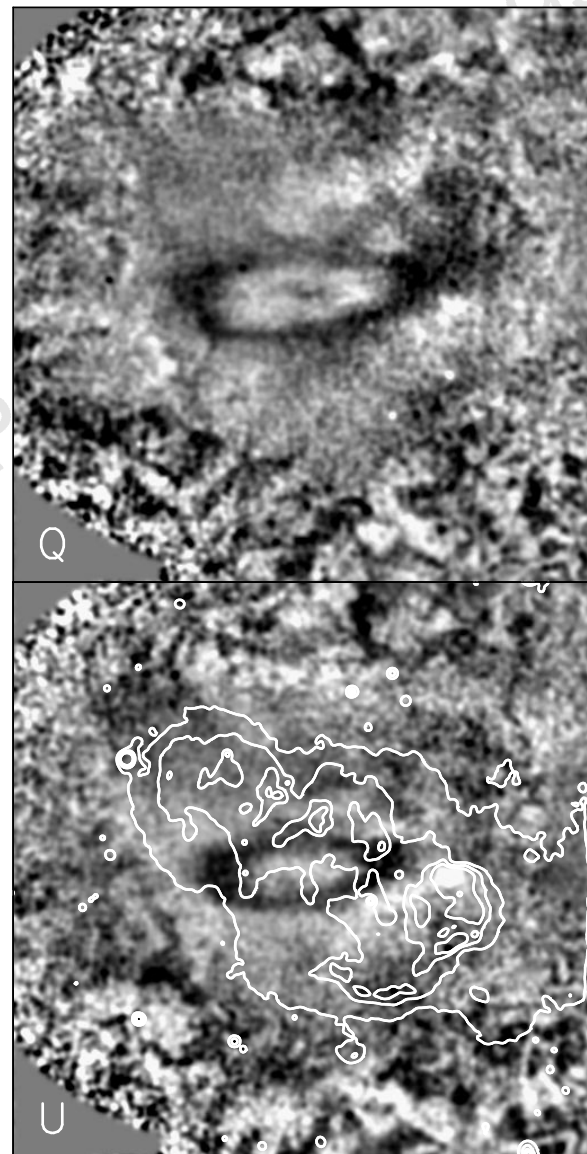


Figure 2 Grey-scale images of Stokes Q (top) and U (bottom), smoothed to $2'$, corresponding to the area shown in Fig. 1. The grey-scale here represents brightness temperature, ranging from -150 mK (black) to $+150$ mK (white). The 1σ noise is ~ 25 mK (the featureless areas at upper- and lower-left are regions of no data). Stokes Q and U parameters are the rectangular components of the linear polarization vectors. The alternating rings of light and dark seen in both images show that Q and U vary in a cyclical manner with radius from the centre of the FRF. This indicates a smooth, linear change of polarization angle. It is these variations of Q and U on spatial scales to which our instrument is sensitive that allow us to detect this large-scale phenomenon. Overlaid on the lower panel are contours of the Stokes I data⁴, also smoothed to $2'$, with contours at 2, 4, 6 and 8 K. These contours clearly show the outline of the large H II region, W5. We note the generally chaotic polarized structures outside W5, and their absence along sight-lines through it. This is a result of strong depolarization effects due to the high rotation measures through the H II region (see text).

the mass becomes very large. On the other hand, for small f the emission measure becomes very high. At sufficiently small f the emission measure should be large enough to produce detectable free-free emission (the intensity of thermal free-free emission from an ionized medium is proportional to emission measure²). However, the strong background emission from W5 (ref. 7) would mask this emission unless the FRF is within a few tens of parsecs from the Sun. The lack of such excess emission implies that the FRF cannot be a local phenomenon.

Limits on f can be obtained by accounting for the polarized emission along the line of sight to W5. The total polarized emission detected in single-antenna measurements²⁰ is 0.5 K; this is a lower limit because small-scale variations in polarization angle will average out within a resolution element. An upper limit of 2 K is obtained by assuming that the total diffuse Galactic emission in this direction²¹ is 50% polarized. Continuum absorption measurements at 22 MHz indicate that ~70% of the total emission lies in front of the W3/W45/W5 region²². We detect 0.23 K of polarized emission interferometrically, which must arise between the FRF and W5, as W5 depolarizes any emission from behind it (any emission originating in front of the FRF must be too smoothly distributed to detect with interferometer as its presence is not seen in our data). Assuming that the total polarized emission is distributed uniformly along the line of sight then implies $0.2 < f < 0.7$. However, in reality the synchrotron emissivity (and hence the polarized emission) of the Galaxy is not uniform, but is concentrated in the spiral arms^{16,17}. In this case the FRF could lie within the local spiral arm at a distance of a few hundred parsecs ($f < 0.2$). Although the free-free emission for this region would remain undetectable against W5, the source of ionizing photons should be visible at this distance.

A further constraint on f might be found with H α /Br γ spectroscopy to search for a velocity component corresponding to the magneto-ionic medium responsible for the FRF. However, the low emission measure, especially seen against the bright emission of W5, and small radial velocity dispersion towards the Perseus arm, only ~50 km s⁻¹, might preclude such a detection.

A search of the SIMBAD database²³ yielded 17 objects within 15' of the FRF's centre, the closest being HD237023, an F4V star only 1.5' away at a distance of <100 pc (based on the distance modulus). However, this star has insignificant ionizing flux and is not powerful enough to have created a magnetic disturbance in the interstellar medium. An O8V star (HD237019) lies 10.7' from the centre, but is part of the OB association that excites W5 (OCl364), and is thus also unlikely to have played a role in the generation of the FRF. In addition to the SIMBAD list, we have identified 8 compact radio sources in our data. The closest is 2.9' from the centre, but no data exist for this or the other radio sources at other frequencies, so we cannot derive radio spectra to aid in identification.

We now consider if the FRF is consistent with an ionized stellar wind. A steady-state stellar wind would produce $n_e \propto r^{-2}$ (centrally peaked), where r is the radius from the central object, and thus a rapid nonlinear change of rotation measure from the centre to the edge of the FRF would be expected. However, we see a linear change of polarization angle from the centre to the observable edge of the FRF (Fig. 1). The absence of a central object could indicate an event-based origin, perhaps the sudden turn-off or removal of a sustaining central object. If so, we can constrain its age. First, if a source of ionizing photons were suddenly removed, then recombination of electrons and ions would occur after $\sim 2 \times 10^6$ yr, a few times the recombination time, $t_r = (\alpha^{(2)} n_e)^{-1}$, where $\alpha^{(2)}$ is the recombination coefficient to the second level of the hydrogen atom. Second, if a source of electrons were suddenly removed, then the dissipation time of the FRF would be a few sound-crossing times²³, $\sim 8f \times 10^6$ yr, assuming 10 km s⁻¹ as the speed of sound.

We conclude that the FRF probably lies 0.2–0.7 times the distance to W5—that is, it is probably an inter-arm feature in the Milky Way—and that it could be a remnant of an event $(2\text{--}6) \times 10^6$ yr ago.

Its lack of fine structure indicates that the magnetic field and electron density are organized smoothly within it. We do not know whether it is caused by the influence of a central object, but if one exists then the observations place constraints on its nature: it must provide electrons with a roughly constant density profile, and at the same time the magnetic field must possess a smooth, non-radial component.

Current theories of the interstellar medium are still quite primitive, predicting structure only on scales the size of whole spiral arms, while acknowledging structure on the scale of the FRF only as a vaguely chaotic. Point sources of energy (for example, supernova remnants, H II regions, and stellar-wind bubbles from massive stars) do generate roughly spherical structures on such scales. However, the FRF apparently contains no central object. The organizing influence is currently a mystery that may be solved by further investigation, possibly deep observations in the optical or infrared wavebands. If there is no central object, then alternative explanations invoking peculiar magnetic structures must be entertained.

Is this a widespread phenomenon? In general, Faraday-rotation observations are subject to confusion if many Faraday screens are superimposed along the line of sight, but the FRF was rendered easily visible due to its alignment with W5. In our observations of the W3/W4/W5 region there are two large H II regions, W4 and W5. Of these two, one showed the phenomenon. Faraday-rotation structure over a large area in the plane of the Milky Way will be imaged in the Canadian Galactic Plane Survey, including a sample of H II regions similar to W3 and W4. If the phenomenon is widespread then other examples should be detected. □

Received 4 June 1997; accepted 17 March 1998.

- Jura, M. in *Interstellar Processes* (eds Hollenbach, D. J. & Thronson, H. A. Jr) 3–17 (Reidel, Dordrecht, 1987).
- Lang, K. R. *Astrophysical Formulae* (Springer, Berlin, 1974).
- Wieringa, M. H., de Bruyn, A. G., Jansen, D., Brouw, W. N. & Katgert, P. Small scale polarization structure in the diffuse galactic emission at 325 MHz. *Astron. Astrophys.* **268**, 215–229 (1993).
- Normandeau, M., Taylor, A. R. & Dewdney, P. E. The Dominion Radio Astrophysical Observatory Galactic Plane Survey pilot project: the W3/W4/W5/HB3 region. *Astrophys. J. Suppl.* **108**, 279–299 (1997).
- Smegal, R. J., Landecker, T. L., Vaneldik, J. F., Routledge, D. & Dewdney, P. E. Aperture synthesis polarimetry: application to the Dominion Radio Astrophysical Observatory synthesis telescope. *Radio Sci.* **32**, 643–656 (1997).
- Salter, C. J. & Brown, R. L. in *Galactic and Extragalactic Radio Astronomy* 2nd edn (eds Verschuur, G. L. & Kellermann, K. I.) 1–36 (Springer, New York, 1988).
- Westerhout, G. A survey of the continuous radiation from the Galactic system at a frequency of 1390 Mc/s. *Bull. Astron. Inst. Neth.* **14**, 215–260 (1958).
- Gardner, F. F. & Whiteoak, J. B. Polarization of cosmic radio waves. *Annu. Rev. Astron. Astrophys.* **4**, 245–292 (1966).
- Braunsfurth, E. Neutral hydrogen in the Cas OB6 association. *Astron. Astrophys.* **117**, 297–303 (1983).
- Broten, N. W., McLeod, J. M. & Vallée, J. P. Catalogue of unambiguous (Faraday-thin, one-component, spectrum-selected) rotation measures for galaxies and quasars. *Astrophys. Space Sci.* **141**, 303–331 (1988).
- Vallée, J. P. in *Fundamentals of Cosmic Physics* Vol. 19, 1–89 (Gordon & Breach, Amsterdam, 1988).
- Taylor, J. H. & Cordes, J. M. Pulsar distances and the Galactic distribution of free electrons. *Astrophys. J.* **411**, 674–684 (1993).
- Rand, R. J. & Kulkarni, S. R. The local Galactic magnetic field. *Astrophys. J.* **343**, 760–772 (1989).
- Beck, R. & Hoernes, P. Magnetic spiral arms in the galaxy NGC6946. *Nature* **379**, 47–49 (1996).
- Fan, Z. & Lou, Y.-Q. Origin of the magnetic spiral arms in the galaxy NGC4946. *Nature* **383**, 800–802 (1996).
- Wilkinson, A. & Smith, F. G. Characteristics of the local Galactic magnetic field determined from background polarization surveys. *Mon. Not. R. Astron. Soc.* **167**, 593–611 (1974).
- Beuermann, K., Kanbach, G. & Berkhuijsen, E. M. Radio structure of the Galaxy: thick disk and thin disk at 408 MHz. *Astron. Astrophys.* **153**, 17–34 (1985).
- Digel, S. W., Lyder, D. A., Philbrick, A. J., Puche, D. & Thaddeus, P. A large-scale CO survey towards W3, W4, and W5. *Astrophys. J.* **458**, 561–575 (1996).
- Leahy, J. P. Small-scale variations in Galactic Faraday rotation. *Mon. Not. R. Astron. Soc.* **226**, 433–446 (1987).
- Brouw, W. N. & Spoelstra, T. A. Th. Linear polarization of the Galactic radio emission at frequencies between 408 and 1411 MHz. Reductions. *Astron. Astrophys. Suppl. Ser.* **26**, 129–146 (1976).
- Kallas, E. & Reich, W. A 21 cm radio continuum survey of the Galactic Plane between $l = 93^\circ$ and $l = 162^\circ$. *Astron. Astrophys. Suppl. Ser.* **42**, 227–243 (1980).
- Roger, R. S. Low frequency absorption due to IC1805 and IC1848. *Astrophys. J.* **155**, 831–840 (1969).
- Dubois, P. in *Library and Information Services in Astronomy* (eds Wilkins, G. A. & Stevens-Rayburn, S.) 146 (US Naval Observatory, Washington DC, 1988).
- Dyson, J. E. & Williams, D. A. *Physics of the Interstellar Medium* 173 (Wiley, New York, 1980).

Acknowledgements. We thank M. Normandeau (Univ. Calgary; now Univ. California at Berkeley) for the acquisition and initial processing of the data, and A. G. Willis (DRAO) for some later processing. The Dominion Radio Astrophysical Observatory is operated as a national facility by the National Research Council of Canada. The Canadian Galactic Plane Survey is a Canadian project with international partners, and is supported by the Natural Sciences and Engineering Research Council of Canada.

Correspondence and requests for materials should be addressed to A.D.G. (e-mail: Andrew.Gray@hia.nrc.ca).

Long-range electrostatic attraction between like-charge spheres in a charged pore

W. Richard Bowen & Adel O. Sharif

Biochemical Engineering Group, Centre for Complex Fluids Processing,
Department of Chemical and Biological Process Engineering, University of Wales
Swansea, Singleton Park, Swansea SA2 8PP, UK

The existence of long-range attractive electrostatic forces between particles of like charge is one of the great current controversies of colloid science. The established theory (Derjaguin–Landau–Verwey–Overbeek; DLVO) of colloidal interactions predicts that an isolated pair of like-charged colloidal spheres in an electrolyte should experience a purely repulsive screened electrostatic (coulombic) interaction^{1,2}. Direct measurements of such interactions have shown quantitative agreement with DLVO theory^{3–5}. Recent experiments, however, provide evidence that the effective interparticle potential can have a long-range attractive component in more concentrated suspensions^{6,7} and for particles confined by charged glass walls^{3,5,8–10}. It is apparent that the long-range attraction in concentrated systems is due to multi-body interactions and may have a similar explanation to the attraction observed for otherwise confined colloids. Theoretical explanations have been proposed^{11–13} but remain the subject of controversy^{14,15}. Here we present a quantitative theoretical explanation of these attractive forces between confined colloidal particles, based on direct solutions of the nonlinear Poisson–Boltzmann equation for two like-charged spheres confined in a cylindrical charged pore. The calculations show that the attraction may be explained by the redistribution of the electric double layers of ions and counterions in solution around the spheres, owing to the presence of the wall; there is thus no need to revise the established concepts underlying theories of colloidal interactions.

The calculation of the interaction between two spheres close to a planar wall, which has been measured experimentally^{6,10}, requires a full three-dimensional solution of the governing equations. However, consideration of the case of like-charge spheres confined in a long, charged cylindrical pore simplifies the calculations by reducing the dimensions of the problem to two. Nevertheless, the calculations remain directly relevant to the experimental situation and have been carried out for conditions of sphere-radius/pore-radius (sphere-radius/sphere-wall separation) ratio, sphere surface potential, pore surface potential and Debye length comparable to those occurring experimentally^{6,10}. We have solved the nonlinear Poisson–Boltzmann equation using an adaptive finite-element method with error minimization¹⁶, and have then calculated the electrostatic forces acting on the spheres.

The normalized Poisson–Boltzmann equation for a two-dimensional electrostatic double layer in cylindrical coordinates may be written as

$$\frac{\partial^2 \Psi}{\partial R^2} + \frac{1}{R} \frac{\partial \Psi}{\partial R} + \frac{\partial^2 \Psi}{\partial Z^2} = -\sigma \quad (1)$$

where the non-dimensional space-charge density is written

$$\sigma = \frac{1}{2} \sum_i z_i C_i \exp(-z_i \Psi) \quad (2)$$

with $C_i = c_i/I$ being the relative concentration. Other variables are defined in Methods. The summation is taken over all ionic species present in the electrolyte. For a symmetric 1:1 electrolyte system, equation (2) reduces to $\sigma = \sinh \Psi$. The dimensionless parameters

and coordinates are defined in accordance with the usual practice in colloid science¹⁶, $R = \kappa r$; $Z = \kappa z$; $\Psi = e\psi/kT$ and

$$\kappa = \left(\frac{e^2 \sum_i z_i^2 c_i}{\epsilon kT} \right)^{1/2} = \left(\frac{2e^2 N_A I}{\epsilon kT} \right)^{1/2}$$

where κ is the Debye parameter and $1/\kappa$ provides a measure of the range of the electrostatic interactions.

Equation (1) is solved with appropriate boundary conditions for the geometry in question. The non-dimensional parameters of relevance are the reduced potential Ψ , the scaled particle radius κa , the sphere-radius to pore-radius ratio λ , and the scaled distance between sphere centres, D ($D = \kappa d$). The electrostatic force acting on the sphere can be obtained directly by integrating the stress tensor, over the sphere surface¹⁷

$$F_E = \int_s \left[\left(\Pi + \frac{1}{2} \epsilon E^2 \right) \mathbf{I} - \epsilon \mathbf{E} \mathbf{E} \right] \cdot \mathbf{n} ds$$

where $\mathbf{E} = -\nabla \psi$ is the electric field and $ds = a^2 \sin \theta d\theta d\phi$. Owing to symmetry, this becomes

$$F_E = \epsilon \pi (\kappa a)^2 \left(\frac{kT}{e} \right)^2 \int_0^\pi |\nabla \Psi|^2 \cos \theta \sin \theta d\theta$$

The results of the calculations are shown in Fig. 1, which indicates how the force varies with the sphere centre–centre distance for two cases: one where the spheres are isolated with radius $a = 0.325 \mu\text{m}$, $\kappa a = 1.185$ and a surface potential $\Psi_s = 3.0$ (top curve); and the corresponding case in which the spheres are confined within a charged pore of surface potential $\Psi_p = 5.0$ and size ratio $\lambda = 0.13$ (bottom curve). These conditions parallel those used in previous experimental work^{6,10} for interacting spheres close to a charged planar wall.

Curve a in Fig. 1 shows that a purely repulsive interaction occurs at all distances for the isolated spheres. This is the interaction used in DLVO theory. Curve b shows that the situation is different for confined spheres of identical surface potential and with identical solution conditions. In this case, the repulsive force at shorter

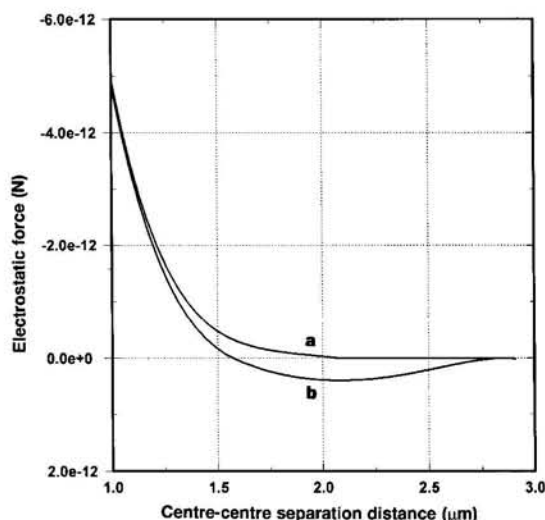


Figure 1 Electrostatic force (F_E) as a function of the sphere centre–centre distance (d) for two cases. **a**, Case where the spheres are isolated with $a = 0.325 \mu\text{m}$, $\kappa a = 1.185$ and a surface potential $\Psi_s = 3.0$. **b**, The corresponding case where the spheres are confined within a charged pore of surface potential $\Psi_p = 5.0$ and size ratio $\lambda = 0.13$. These conditions parallel those in previous experimental work^{6,10} for interacting spheres close to a charged planar wall.

sphere–sphere separation distances is essentially the same as that for the isolated spheres. However, as the separation increases, so the repulsive force becomes less than that for the isolated spheres. Eventually, at even greater distances, the force changes sign, that is it becomes attractive. At very large distances, the force approaches zero. The attractive minimum is similar in form to those previously reported experimentally^{6,10}. The magnitude of the attractive interaction potential corresponding to the force–distance curve in Fig. 1 is about an order of magnitude greater than the experimental value^{6,10} for a planar surface due to the different geometry in the calculations (cylindrical) and possibly some uncertainty in the experimental surface potential. It is, however, of the order expected for interactions in concentrated colloidal dispersions²⁰, for which the cylindrical geometry is a useful analogy. It should be noted that the attractive minimum arises naturally when proper account is taken of the presence of the wall in the calculations. It is not necessary to introduce any further assumptions, and the explanation is based on established colloid theory.

The attractive force arises as a result of a redistribution of the static electrical double layers around the spheres caused by the presence of the wall. This is apparent in Fig. 2, where isopotential lines are shown for the problem considered. Each part of the figure shows a half-section which, because of symmetry, is sufficient to describe the physical reality. (Numerical computation also makes use of the plane of symmetry between the spheres.) Figure 2a shows the calculated potential distribution for the isolated spheres; Fig. 2b shows the calculated potential distribution for the spheres confined in a pore. Comparison shows that in the latter case the potential distribution is disturbed both between the spheres and on opposing sides of the spheres. It is this modified distribution that results in a net attractive force due to the presence of the pore wall. This modified distribution is shown quantitatively in Fig. 3, which shows the potential along the midplane between the spheres for the two cases. In the absence of the wall, the potential along this plane is low at all distances. However, the presence of the wall induces a significant potential between the spheres (Fig. 3, inset).

Our calculations confirm speculation that the attraction between confined spheres may result from the electrostatic influence of the confining plates⁸. Further, the case of two interacting spheres in a tube could be taken as a very first approximation to multiparticle interactions in concentrated systems. The attractive minima calcu-

lated here could therefore provide the basis of an explanation for phase separation and the formation of ordered structures in some concentrated colloidal systems^{6,7}. On this basis, curve b of Fig. 1 also shows that the repulsive force between particles in concentrated systems is expected to be less than that for isolated pairs of spheres, even for separations at which no attractive force exists, as has been found experimentally. The assessment provided here is in agreement with previous qualitative explanations of phase separation¹⁸. Other workers have also shown theoretically that a purely repulsive screened-Coulomb (DLVO) interaction between charged colloidal spheres is compatible with phase separation at low ionic strength¹⁹. It is not necessary to introduce a new description of interaction, such as a Sogami–Ise potential, as has been suggested²⁰. Indeed, the Sogami–Ise approach^{11,12} indicates a long-range attractive potential even for a pair of isolated spheres, which we do not find.

Our calculations indicate that further experimental and theoretical investigations of these phenomena could be very rewarding, for long-range attraction may have an important influence on the properties of colloids of industrial significance. Our results also show that the existence of such attraction does not in itself lead to a requirement to reformulate the basis of existing theories of colloidal systems. □

Methods

Variables. a , particle radius (m); c_i , concentration of ion species i (kmol m⁻³); C_i , relative concentration of ion species (c_i/I); d , centre–centre separation distance (m); D , reduced centre–centre distance; e , electronic charge (1.602×10^{-19} C); E , electric field (V m⁻¹); F_E , electrostatic force (N); I , ionic strength of the electrolyte, in kmol m⁻³ ($= (\sum z_i^2 c_i)/2$); $\mathbf{I}$, unit tensor; k , Boltzmann constant (1.381×10^{-23} J K⁻¹); $\mathbf{n}$, outward normal; N_A , Avogadro's number (6.022×10^{26} kmol⁻¹); r , radial coordinate (m); R , dimensionless coordinates ($R = \kappa r$); s , particle surface area (m²); T , absolute temperature (K); z , axial coordinate (m); Z , dimensionless coordinates ($Z = \kappa z$); z_i , charge number of ion species (i); ϵ , permittivity of the electrolyte ($= \epsilon_0 \epsilon_r$), in F m⁻¹; ϵ_0 , permittivity of vacuum (8.854×10^{-12} in F m⁻¹); ϵ_r , relative permittivity (dielectric constant); κ , Debye parameter (m⁻¹); θ , angle subtended at a centre of circle by an arc (rad); λ , sphere–radius/pore–radius ratio; Π , difference in local osmotic pressure (N m⁻²); σ , non-dimensional space charge density; ϕ , azimuth angle in a plane normal to the axis (rad); ψ , electric potential (V); Ψ , reduced potential; Ψ_s , reduced potential at sphere surface; Ψ_p , reduced potential at pore wall.

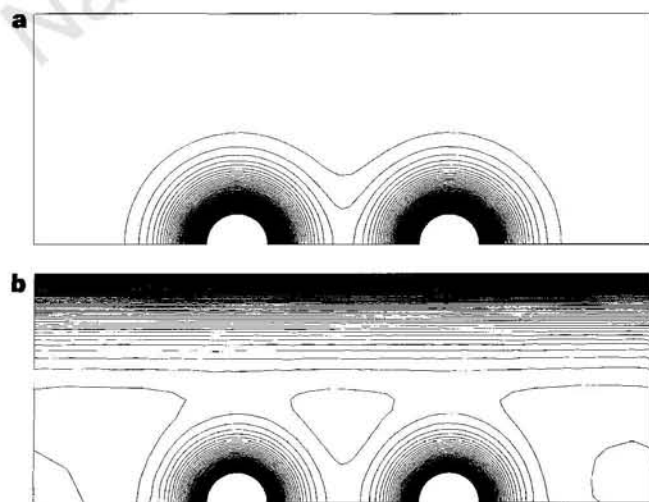


Figure 2 Calculated isopotential lines for the problem considered. A half-section of the physical geometry is shown, with the line of symmetry lying at the bottom. **a**, Isolated spheres; **b**, spheres confined in a pore. The pore wall is at the top.

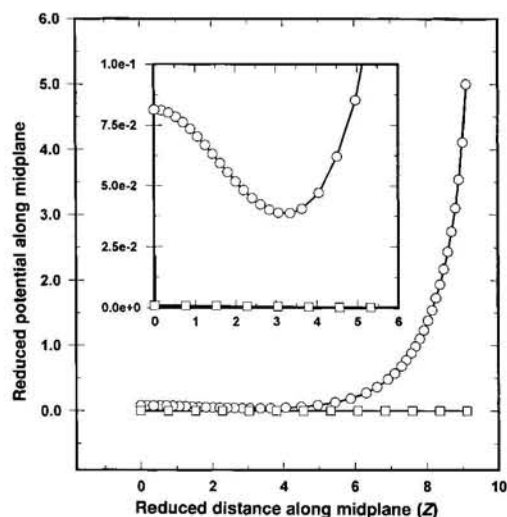


Figure 3 Potential along the midplane between two spheres: isolated spheres (square symbols) and spheres confined in a pore (circles). Zero distance corresponds to the line of centres of the spheres. On the right-hand side of the main graph, the pore wall is reached at $\Psi = 5.0$.

Received 19 November 1997; accepted 31 March 1998.

- Derjaguin, B. V. & Landau, L. Theory of the stability of strongly charged lyophobic sols and the adhesion of strongly charged particles in solutions of electrolytes. *Acta Physicochimica* (USSR) **14**, 633 (1941).
- Verwey, E. J. & Overbeek, J. Th. G. *Theory of the Stability of Lyophobic Colloids* (Elsevier, Amsterdam, 1948).
- Crocker, J. C. & Grier, D. G. Microscopic measurements of pair interaction potential of charge-stabilised colloid. *Phys. Rev. Lett.* **73**, 352–355 (1994).
- Crocker, J. C. & Grier, D. G. Methods of digital video microscopy for colloidal studies. *J. Coll. Interface Sci.* **179**, 298–310 (1996).
- Grier, D. G. Optical tweezers in colloid and interface science. *Curr. Opin. Coll. Interface Sci.* **2**, 264–270 (1997).
- Larsen, A. E. & Grier, D. G. Like-charge attractions in metastable colloidal crystallites. *Nature* **385**, 230–233 (1997).
- Ise, N., Okubo, T., Sugimura, M., Ito, K. & Nolte, H. J. Ordered structure in dilute solutions of highly charged polymers lattices as studied by microscopy. I. Interparticle distance as a function of latex concentration. *J. Chem. Phys.* **78**, 536–540 (1983).
- Kepler, G. M. & Fraden, S. Attractive potential between confined colloids at low ionic strength. *Phys. Rev. Lett.* **73**, 356–359 (1994).
- Carbajal-Tinoco, M. D., Castro-Roman, F. & Arauz-Lara, J. L. Static properties of confined colloidal suspensions. *Phys. Rev. E* **53**, 3745–3749 (1996).
- Crocker, J. C. & Grier, D. G. When like charge particles attract: the effects of geometrical confinement on long-range colloidal interactions. *Phys. Rev. Lett.* **77**, 1897–1900 (1996).
- Sogami, I. Effect potential between charged spherical particle in dilute suspension. *Phys. Lett. A* **96**, 199–203 (1983).
- Sogami, I. & Ise, N. On the electrostatic interaction in macroionic solutions. *J. Chem. Phys.* **81**, 6320–6332 (1984).
- Chu, X. & Wasan, D. Attractive interaction between similarly charged colloidal particles. *J. Coll. Interface. Sci.* **184**, 268–278 (1996).
- Jönsson, B., Åkesson, T. & Woodward, C. E. in *Ordering and Phase Transitions in Colloidal Systems* (eds Arora, A. K. & Tata, B. V. R.) Ch. 11, 295–313 (VCH, New York, 1996).
- Smalley, M. V. in *Ordering and Phase Transitions in Colloidal Systems* (eds Arora, A. K. & Tata, B. V. R.) Ch. 12, 315–337 (VCH, New York, 1996).
- Bowen, W. R. & Sharif, A. O. Adaptive finite element solution of the non-linear Poisson–Boltzmann equation: a charged spherical particle at various distances from a charged cylindrical pore in a charged planar surface. *J. Coll. Interface Sci.* **187**, 363–374 (1997).
- Russel, W. B., Saville, D. A. & Schowalter, W. R. *Colloidal Dispersions* (Cambridge Univ. Press, 1989).
- Hastings, R. On the crystallisation of macroionic solutions. *J. Chem. Phys.* **68**, 675–678 (1978).
- van Roij, R. & Hansen, J.-P. Van der Waals-like instability is suspensions of mutually repelling charged colloids. *Phys. Rev. Lett.* **79**, 3082–3085 (1997).
- Tata, B. V. R., Yamahara, E., Rajamani, P. V. & Ise, N. Amorphous clustering in highly charged dilute poly(chlorostyrene-styrene sulfonate) colloids. *Phys. Rev. Lett.* **78**, 2660–2663 (1997).

Acknowledgements. This work was funded by the UK Biotechnology and Biological Sciences Research Council.

Correspondence should be addressed to W.R.B. (e-mail: r.bowen@swansea.ac.uk).

Non-volatile holographic storage in doubly doped lithium niobate crystals

K. Buse, A. Adibi & D. Psaltis

California Institute of Technology, Department of Electrical Engineering, Pasadena, California 91125, USA

Photorefractive materials are being widely investigated for applications in holographic data storage¹. Inhomogeneous illumination of these materials with an optical interference pattern redistributes charge, builds up internal electric fields and so changes the refractive index. Subsequent homogeneous illumination results in light diffraction and reconstructs the information encoded in the original interference pattern. A range of inorganic and organic photorefractive materials are known², in which thousands of holograms of high fidelity can be efficiently stored, reconstructed and erased. But there remains a problem with volatility: the read-out process usually erases the stored information and amplifies the scattered light. Several techniques for ‘fixing’ holograms have been developed^{3–6}, but they have practical disadvantages and only laboratory demonstrators have been built^{7–10}. Here we describe a resolution to the problem of volatility that should lead to the realization of a more practical system. We use crystals of lithium niobate—available both in large size and with excellent homogeneity—that have been doped with two different deep electron traps (iron and manganese). Illumination of the crystals with incoherent ultraviolet light during the record-

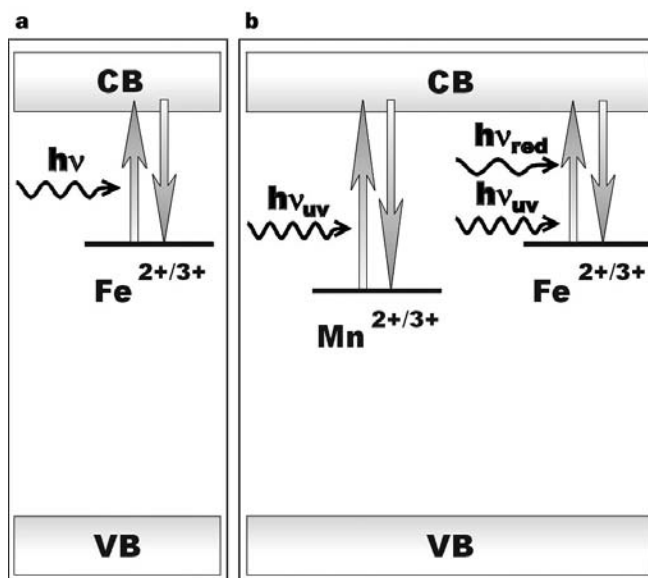


Figure 1 Band diagram of LiNbO₃ doped with iron (a) and doubly doped with iron and manganese (b). CB, conduction band; VB, valence band. Manganese has a deeper energy level than iron. Both elements occur in two different valence states. Electrons are excited into the conduction band from Fe²⁺ either by visible light or by ultraviolet light, or from Mn²⁺ by ultraviolet light only. Conduction-band electrons can recombine with Fe³⁺ or with Mn³⁺.

ing process permits the storage of data (a red-light interference pattern) that can be subsequently read, in the absence of ultraviolet light, without erasure. Our crystals show up to 32 per cent diffraction efficiency, rapid optical erasure of the stored data is possible using ultraviolet light, and light scattering is effectively prevented.

Build-up of space-charge fields in photorefractive materials requires redistribution of charge. Transition-metal ions can occur in inorganic crystals in different valence states: for example, iron is present in LiNbO₃ as Fe²⁺ and Fe³⁺ (ref. 11). The energy level of such a centre is located in the gap between valence and conduction band; Fig. 1a illustrates this. Iron is a deep centre, and hence thermal excitations are negligible. Light, however, excites electrons from Fe²⁺, the electrons migrate in the conduction band and finally they are trapped somewhere else by Fe³⁺ sites. The dominant charge driving force in LiNbO₃ is the bulk photovoltaic effect, which causes the excited electrons to migrate in a preferred direction. Drift and diffusion can also contribute to the movement of the charge carriers. Reconstruction of stored information, which requires homogeneous illumination with light of the recording wavelength, causes uniform electron excitation. As a result, the charge pattern is ‘washed out’ and the hologram becomes erased.

One approach to obtaining non-volatile storage is to copy the space-charge pattern into a pattern of immobile ions³ (‘thermal fixing’). Ions become mobile if the crystals are heated; they migrate to compensate the space-charge field and after cooling they become immobile again. The resulting ion pattern cannot be erased by light. Alternatively, the recorded space-charge field can be copied into ferroelectric domains⁴ (‘electrical fixing’). This is accomplished through application of external electric fields, which switch ferroelectric domains in the regions where the external and internal fields add up. The domain pattern is also stable against further illumination. Unfortunately, heating or application of large external fields is not practical and rapid refreshing of the memory is not possible. Two-step recording⁵ is an all-optical approach for non-volatile storage: recording light of low photon energy and sensitizing light

of high photon energy excite electrons via virtual or real intermediate levels to the conduction band. For read-out only, light of low photon energy is used which cannot by itself generate free electrons, and thus the stored charge pattern is stable. Despite impressive recent progress¹², the low sensitivity of this approach remains a serious drawback. Recording with light of one wavelength and just reading with light of a longer one causes substantial information losses due to the Bragg condition of volume diffraction⁶. Even though all these techniques work^{7–10}, the outlined disadvantages have made holographic data storage impractical thus far.

The method we describe here uses crystals that have two different deep traps, with energy levels in the gap between the valence and conduction bands. Such crystals are often photochromic^{13,14}, which means that the absorption of the material at one wavelength can be changed by illuminating it at another wavelength. We can explain the physical mechanism for photochromism in such doubly doped crystals with reference to Fig. 1b, which is the band diagram for a LiNbO₃ crystal doped with iron (shallower trap) and manganese (deeper trap). Manganese occurs in LiNbO₃ in the valence states 2+ and 3+ (ref. 15). Initially, the electrons tend to be in the deeper traps and the crystal is transparent for wavelengths corresponding to the Fe^{2+/3+} level (centred at 477 nm). If the crystal is illuminated with light at shorter wavelengths that can ionize the deeper manganese traps, then the iron traps become populated and the crystal becomes absorptive for a wide range of visible wavelengths. We can return the crystal to its transparent state by illuminating it with light in the visible range, which transfers the electrons from the iron traps back to the manganese traps. This is shown in Fig. 2, which is a photograph of such a crystal after ultraviolet pre-exposure and subsequent red-light illumination. The red light was present in the centre of the crystal only, and the bleaching can be clearly seen. It has been previously observed that the holographic sensitivity of such photochromic photorefractive crystals is enhanced for longer wavelengths when they are pre-exposed to short-wavelength radiation^{13,16}. Here we show how we can use photochromic



Figure 2 Photo-assisted electron transfer between traps in doubly doped LiNbO₃. Photograph of a LiNbO₃ crystal (dimensions 5 × 4 × 0.85 mm³) doped with 0.075 wt% Fe₂O₃ and 0.01 wt% MnO after homogeneous illumination with ultraviolet light (wavelength 365 nm, intensity 20 mW cm⁻², 120 min) and subsequent illumination of the middle of the crystal face with a laser light beam (wavelength 633 nm, intensity 2,400 mW cm⁻², 1/e² diameter 1 mm, 30 min). The ultraviolet light excites electrons from manganese into iron, creating Fe²⁺. This centre absorbs visible light and the crystal is darkened. The red light excites the electrons from iron and they are trapped in manganese, where they cannot be excited by visible light. Thus the crystal is bleached by the red light.

doubly doped LiNbO₃ crystals to realize non-volatile holographic storage.

Experiments were performed with a 0.85-mm-thick LiNbO₃ crystal doped with 0.075 wt% Fe₂O₃ and 0.01 wt% MnO. We used a 100-W mercury lamp as the ultraviolet light source, and a 35-mW HeNe laser for holographic recording. The unpolarized ultraviolet light illuminates the crystal evenly (wavelength 365 nm, intensity 20 mW cm⁻²); the HeNe laser light is split into two plane waves which interfere at the crystal (wavelength 633 nm, beam diameter 2.0 mm, intensity of each wave 300 mW cm⁻², transmission geometry; quoted diameter and intensity are for the area within which power drops to 1/e² of its maximum value). The grating vector of the interference pattern is aligned along the *c* axis of the sample. During recording, one of the HeNe beams is blocked from time to time and the second beam is diffracted from the written grating to obtain the diffraction efficiency η as the ratio between diffracted and total incident light powers.

The lower curve in Fig. 3 shows the evolution of the diffraction efficiency when a holographic grating is recorded with the HeNe laser light beams only, following a 2-hour pre-exposure to ultraviolet light. It increases rapidly, reaches a maximum and decreases afterwards almost to zero. The origin of this behaviour is clear: the ultraviolet pre-exposure excites electrons from manganese centres and populates homogeneously the iron level. As the Fe²⁺ ions can absorb red light, the HeNe laser records a hologram. Interference maxima yield large photovoltaic currents, which build up space-charge fields and refractive-index changes. However, the Fe²⁺ sites become bleached in the high-intensity regions and the current drops. Ultimately, the darker regions are also bleached, and all electrons are trapped by the Mn³⁺ ions. However, the achieved Mn²⁺ concentration is almost completely spatially homogeneous; this is because we start with a homogeneous Fe²⁺ concentration, and each excited carrier is moved in the same direction by approximately the same distance before it becomes retrapped by the Mn³⁺ ions. Thus the final space-charge field is very small, and this approach is not suitable for efficient storage.

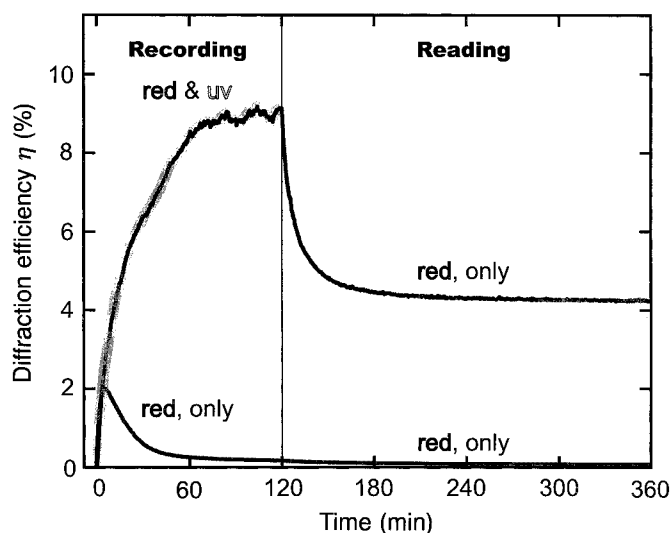


Figure 3 Holographic recording and read-out curves. The diffraction efficiency η , the ratio between the intensity of the diffracted and of the total incident read-out light, is shown against time. Lower curve: the crystal is pre-exposed to the ultraviolet light (wavelength 365 nm, intensity 20 mW cm⁻², 120 min), then the hologram is recorded (wavelength 633 nm, ordinary polarization, intensity 600 mW cm⁻², 120 min) and finally it is read by one of the recording beams (633 nm, intensity 300 mW cm⁻², 240 min). Upper curve: same procedure, but the ultraviolet light is also present during recording. Simultaneous presence of ultraviolet and red light yields higher efficiencies and in the final stage, non-volatile read-out.

The central idea that allows non-volatile storage is to illuminate the crystal with ultraviolet and red light simultaneously, wait until saturation is reached and then switch the ultraviolet light off. The upper curve in Fig. 3 shows the result obtained by this method. The first observation is that much larger efficiencies are achieved during recording. In addition, read-out erases only a fraction of the hologram because electrons are removed from iron. However, after complete bleaching of the Fe^{2+} sites, the hologram remains recorded in the manganese traps. From this stage on, the crystal is insensitive to red light and read-out is non-volatile. Extrapolation of a 12-hour reading experiment shows that continuous read-out of at least two weeks is possible until the diffraction efficiency drops further by a factor of $1/e$.

Figure 4 illustrates the processes schematically. The light intensities and the concentrations of electrons trapped in iron and manganese are shown against the spatial coordinate before, during and after recording. At the beginning, only homogeneous ultraviolet light is present (Fig. 4a). Some electrons are trapped in manganese and some in iron.

During recording (Fig. 4b), the presence of the red interference pattern causes electrons to be excited from iron and moved back to manganese in the regions of interference maxima. The bulk photo-voltaic currents, which are proportional to the product of light

intensity and concentration of filled electron traps, create spatially inhomogeneous currents that contribute to build-up of the space-charge field. Homogeneous ultraviolet light excites electrons from the modulated iron and manganese patterns, while modulated red light excites electrons only from iron. As the excited electron distribution and current pattern are replicas of the red-light intensity pattern, the distribution of the charge re-trapped in manganese also stores the same pattern. The continuous presence of ultraviolet light replenishes the iron sites and makes it possible to sustain and build up the transfer of information from iron to manganese. Detailed analysis of this process yields the following result: saturation space-charge field, refractive-index change and diffraction efficiency depend on the ratio of the intensities of the red and ultraviolet light, and not on their absolute values. Red light of too high an intensity bleaches the Fe^{2+} sites, the concentration of electrons in iron is small, and the space-charge field approaches zero. If the red light is too weak, the ultraviolet light erases the information, and the space-charge field again becomes zero.

After recording, the ultraviolet light is switched off. The red light initially removes the electrons from iron until all of them are trapped in manganese, and read-out becomes non-volatile (Fig. 4c). However, optical erasure by homogeneous ultraviolet light is still possible.

Our experiments reach a non-volatile diffraction efficiency of 4% for ordinarily polarized light and, due to a larger electro-optic coefficient, 32% for extraordinary polarization. The square root of the saturation efficiency yields the quantity $M/\#$, a measure of how many holograms can be multiplexed¹⁷. We get a maximum $M/\#$ of ~ 0.6 with the 0.85-mm-thick crystal. Typical values for volatile recording with green light in iron-doped material are around $M/\# = 1$ for ~ 1 -cm-thick crystals¹⁷. As $M/\#$ is proportional to the thickness of the recording material, a 1-cm-thick crystal with the new technique should yield $M/\# \approx 7$. The areal density of absorbed light energy required to get 1% diffraction efficiency is denoted by $W_{\eta=1\%}$. Our non-volatile method requires $W_{\eta=1\%} = 3.5 \text{ J cm}^{-2}$ of red light, while typical values achieved with green light in iron-doped material are $\sim 1 \text{ J cm}^{-2}$. For instance, for a recording area of 1 mm^2 , and 10 mW of absorbed red light, the exposure time required to achieve diffraction efficiency equal to 10^{-6} is less than 35 ms, corresponding to a recording rate of $\sim 30 \text{ Mbit s}^{-1}$. It is remarkable that the new technique provides a significant improvement in $M/\#$ and does not sacrifice sensitivity ($W_{\eta=1\%}$), even though it was produced in order to provide convenient non-volatile storage and its parameters here not yet been optimized. For instance, the concentrations of iron and manganese can be fine-tuned. Use of cerium instead of iron, and copper instead of manganese, may further improve both $M/\#$ and $W_{\eta=1\%}$. A change of the wavelength of the sensitizing light may enable additional advances to be made.

The main advantages of our technique compared to other non-volatile storage methods are that there is no need for external electric fields or light of high intensity. Furthermore, during recording, the ultraviolet light prevents build-up of both holographically amplified scattered light and screening fields created by accumulation of charge at the boundaries of illuminated regions. During read-out, the crystal is insensitive and cross-talk build-up due to two-wave mixing effects does not occur. Thus the fidelity of the stored information is significantly improved and the error rate drops. Smooth and continuous recording curves like those shown in Fig. 3 are hard to obtain in conventional experiments. It appears that we have discovered a recording method in which essentially all the desirable features of holographic memory are improved. We consider that the prospects are excellent for practical holographic storage devices using this technique. □

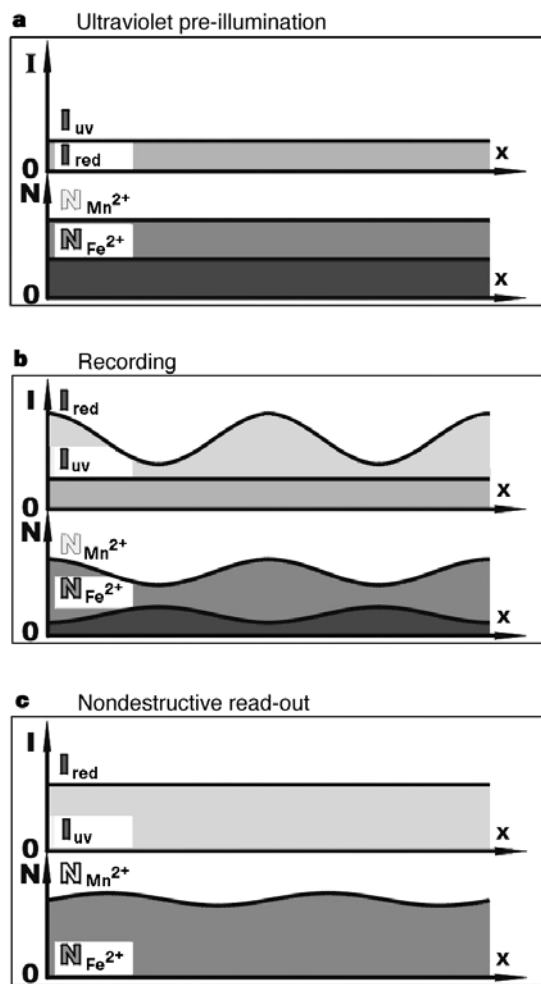


Figure 4 Recording and non-destructive read-out processes. Intensities of red and ultraviolet light (I_{red} and I_{uv}), and concentrations of electrons trapped in iron and manganese ($N_{\text{Fe}^{2+}}$ and $N_{\text{Mn}^{2+}}$) against spatial coordinate x along the grating vector during ultraviolet pre-illumination (a), recording (b) and non-destructive read-out (c).

Received 4 February; accepted 22 April 1998.

1. Psaltis, D. & Mok, E. Holographic memories. *Sci. Am.* **273**, 70–76 (1995).
2. Meerholz, K., Volodin, B. L., Sandalphon, Kippelen, B. & Peyghambarian, N. A photorefractive polymer with high optical gain and diffraction efficiency near 100%. *Nature* **371**, 497–500 (1994).

- Amodei, J. J. & Staebler, D. L. Holographic pattern fixing in electro-optic crystals. *Appl. Phys. Lett.* **18**, 540–542 (1971).
- Micheron, F. & Bismuth, G. Electrical control of fixation and erasure of holographic patterns in ferroelectric materials. *Appl. Phys. Lett.* **20**, 79–81 (1972).
- von der Linde, D., Glass, A. M. & Rodgers, K. F. Multiphoton photorefractive processes for optical storage in LiNbO_3 . *Appl. Phys. Lett.* **25**, 155–157 (1974).
- Külich, H. C. A new approach to read volume holograms at different wavelengths. *Opt. Commun.* **64**, 407–411 (1987).
- Heanue, J. F., Bashaw, M. C., Daiber, A. J., Snyder, R. & Hesselink, L. Digital holographic storage system incorporating thermal fixing in lithium niobate. *Opt. Lett.* **21**, 1615–1617 (1996).
- Ma, J. *et al.* Electrical fixing of 1000 angle-multiplexed holograms in SBN:75 . *Opt. Lett.* **22**, 1116–1118 (1997).
- Lande, D., Orlov, S. S., Akella, A., Hesselink, L. & Neurgaonkar, R. R. Digital holographic storage system incorporating optical fixing. *Opt. Lett.* **22**, 1722–1724 (1997).
- Chuang, E. & Psaltis, D. Storage of 1000 holograms with use of a dual-wavelength method. *Appl. Opt.* **36**, 8445–8454 (1997).
- Kurz, H. *et al.* Photorefractive centers in LiNbO_3 studied by optical, Mössbauer- and EPR-methods. *Appl. Phys. Lett.* **12**, 355–368 (1977).
- Guenther, H., Wittmann, G., Macfarlane, R. M. & Neurgaonkar, R. R. Intensity dependence and white-light gating of two-color photorefractive gratings in LiNbO_3 . *Opt. Lett.* **22**, 1305–1307 (1997).
- Staebler, D. L. & Phillips, W. Hologram storage in photochromic LiNbO_3 . *Appl. Phys. Lett.* **24**, 268–270 (1974).
- Ming, Y., Krätzig, E. & Orlowski, R. Photorefractive effects in $\text{LiNbO}_3:\text{Cr}$ induced by two-step excitations. *Phys. Status Solidi A* **92**, 221–229 (1985).
- Thiemann, O. & Schirmer, O. F. Energy levels of several 3d impurities and EPR of Ti^{3+} in LiNbO_3 . *Proc. SPIE* **1018**, 18–22 (1988).
- Buse, K., Holtmann, L. & Krätzig, E. Activation of BaTiO_3 for infrared holographic recording. *Opt. Commun.* **85**, 183–186 (1991).
- Mok, F. H., Burr, G. W. & Psaltis, D. System metric for holographic memory systems. *Opt. Lett.* **21**, 896–898 (1996).

Acknowledgements. We thank E. Krätzig for discussions. This work was supported by JPL, funded by DARPA/ITO, and by grants from Rome Labs. K.B. thanks the Deutsche Forschungsgemeinschaft for a postdoctoral fellowship.

Correspondence and requests for materials should be addressed to D.P. (e-mail: psaltis@sunoptics.caltech.edu)

Isolation and properties of small-bandgap fullerenes

Michael D. Diener & John M. Alford

TDA Research Inc., 12345 West 52nd Avenue, Wheat Ridge, Colorado 80033, USA

The diversity of molecular structures exhibited by fullerenes¹ suggests a wide range of interesting and useful properties. Several fullerenes are now considered to be well characterized, but only minor variations in their chemical and physical properties have been observed². Here we show that there are in fact two distinct classes of fullerenes, with some very different chemical properties. Members of the first class, typified by C_{60} and C_{70} , have large energy gaps between the highest occupied and lowest unoccupied molecular orbitals (HOMO and LUMO), and are soluble in many organic solvents. The second, previously unrecognized class is represented by C_{74} and selected isomers of the higher fullerenes, such as that of C_{80} with icosahedral symmetry: these are either free radicals or have small HOMO–LUMO gaps. Like radical metallofullerenes, they are kinetically unstable and react readily to form insoluble, polymerized solids. These intermolecular bonds can be broken by electrochemical reduction. After reducing them to soluble anions, we have been able to isolate and characterize these new fullerenes.

The solubility of fullerenes in common organic solvents is a primary factor contributing to the rapid advancement of fullerene chemistry. Solvation allows extraction of fullerene clusters from carbon soot³, isolation of individual structures, and derivatization to form new compounds. Isomers of C_{60} through to C_{96} have been isolated and identified^{4–6}, and these extractable isomers share common electronic features⁷. Specifically, both experimental^{8–10} and computational^{11,12} investigations show that all soluble fullerenes have large bandgaps. As small-bandgap and radical fullerene structures were not identified in the extracts, it was generally concluded that these isomers were thermally annealed into more stable structures and were not present in the soot.

C_{74} is an anomaly because it is regularly observed in soots by mass spectrometry, but it is not soluble. Like C_{60} and C_{70} , C_{74} possesses only one possible isolated pentagon structure. Previous low-level Hückel molecular orbital^{1,11} and tight-binding molecular dynamics calculations^{7,12} on $\text{C}_{74}(\text{D}_3)$ predicted a small bandgap, and it has been suggested that this could explain its insolubility⁷. More accurate density functional theory calculations (Fig. 1) confirm a very small (0.05 eV) bandgap, and thermal population of the triplet state will produce a chemically reactive biradical. Adding two electrons to C_{74} creates a kinetically stable (and soluble) closed-shell configuration with a moderate bandgap. An example is $\text{Ca}@\text{C}_{74}$ (ref. 13), a soluble endohedral metallofullerene in which two electrons have been donated from the Ca atom, filling the LUMO.

Observations are similar for the higher fullerenes. Only a small percentage of the C_{80} found in soots can be solvent-extracted, and this large-bandgap D_2 isomer¹⁴ may not be as thermodynamically stable as the D_{5h} or I_h radical isomers^{12,15}. As the undistorted ground state HOMO of $\text{C}_{80}(\text{I}_h)$ has only two electrons in four orbitals, the addition of six electrons produces a closed-shell, large-bandgap species, explaining the solubility of $\text{La}_2@\text{C}_{80}$ (ref. 16).

It is likely that larger fullerenes ($>\text{C}_{100}$) consist almost entirely of insoluble small-bandgap structures. Calculations indicate that the statistical probability of forming closed-shell, large-bandgap structures past C_{100} is very small¹, and this could account for the large amount of insoluble giant fullerenes observed in the soots. A report¹⁷ demonstrating that giant fullerenes (as well as C_{74} and C_{80}) can be extracted after chemical derivatization to passivate radical structures strongly suggests that they too have radical or small-bandgap configurations. Although we have considered only isolated pentagon isomers with small bandgaps, the soot could also contain adjacent pentagon isomers which are highly reactive owing to excessive pyramidalization (steric strain) along the pentagon junctions³¹.

With small-bandgap (and radical) electronic structures, C_{74} and $\text{C}_{80}(\text{I}_h)$ exhibit properties similar to the endohedral lanthanide metallofullerenes. The majority of radical Ln-metallofullerenes are insoluble, and only certain isomers of $\text{Ln}@\text{C}_{82}$ can be readily extracted¹⁸. The unique solubility of $\text{Ln}@\text{C}_{82}$ remains a mystery, but it is probably due to localization of the radical electron in bonds near the Ln atom. In addition to their similar electronic configurations, the unfilled orbitals of both small-bandgap fullerenes

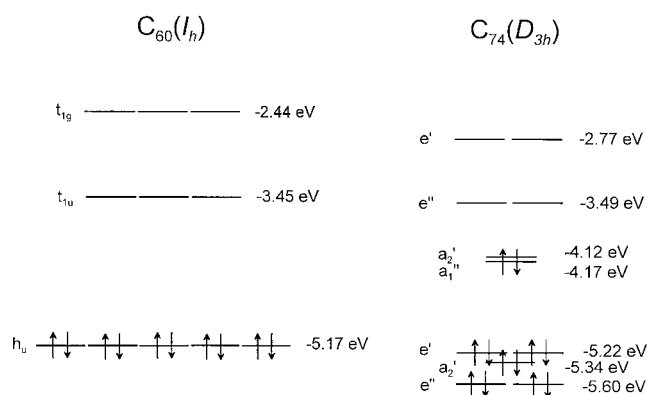


Figure 1 Kohn-Sham orbital energy levels of C_{60} and C_{74} . Density functional theory calculations were performed using the BLYP functional in Gaussian 94 (ref. 28). Geometries were optimized using the 3-21G* basis set, followed by single point calculations using the 6-31G* basis set. The calculated value of 1.72 eV for the C_{60} bandgap is in good agreement with a previous calculation of 1.64 eV (ref. 29). A singlet state was assumed for C_{74} .

and Ln-metallofullerenes are lower in energy than the LUMO of the soluble fullerenes, as verified by gas-phase electron-affinity measurements^{19,20}, electrochemical studies²¹ and calculations (Fig. 1).

Given the low level of these unfilled orbitals, we suggest that both small-bandgap fullerenes and radical metallofullerenes can form insoluble, polymerized solids. The propensity of fullerenes with similar electronic configurations to polymerize has already been demonstrated by $C_{59}N$ dimers²², $Li@C_{60}$ oligomers²³, and conducting polymer chains in A_1C_{60} (where A is K or Rb)²⁴. As the unfilled orbitals move lower in energy, this tendency to polymerize should increase. We show here that electrochemical reduction breaks these intercage bonds, and as anions, the insoluble small-bandgap fullerenes and metallofullerenes can be isolated and characterized.

Fullerenes and Gd-metallofullerenes were produced by the carbon arc method in a quartz bell-jar reactor designed for both production and subsequent anaerobic sublimation of fullerenes from the raw soot. Sublimation is an effective method for separating both small-bandgap fullerenes (such as C_{74})²⁵ and metallofullerenes¹⁸ from graphitic fragments, metal carbides and other non-volatile radicals that could interfere with the electrochemical processing. After arcing, the reactor was heated slowly to 750 °C, and the resulting fullerene sublimate was collected on a water-cooled cold finger. (For further details of the separation process, see Supplementary Information.) Laser desorption mass spectrometry (Figs 2a, 3a) shows that fullerene and metallofullerene species up to $\sim C_{100}$ sublimed. All further processing was performed in a helium-atmosphere glovebox.

The sublimates were thoroughly washed with *o*-xylene to remove soluble species. Approximately 20% of the metallofullerene sublimate and 10% of the empty fullerene sublimate did not dissolve. Analysis of the *o*-xylene solutions (Figs 2b, 3b) indicates that C_{74} , C_{80} , many higher fullerenes, and most Gd-metallofullerenes are insoluble. Analysis of the washed, insoluble sublimate showed that substantial amounts of soluble C_{60} and C_{70} remained trapped in the particulates.

The insoluble particulates were suspended by sonication in benzonitrile (~ 25 mg in 25 ml) containing 0.1 M (*n*-Bu)₄NPF₆ as supporting electrolyte and electrochemically reduced using a technique similar to that used for bulk reduction of C_{60} in acetonitrile²⁶. The potential was held constant at -1.0 V relative to a Ag/AgNO₃ reference. During reduction, both the empty and metallofullerene samples produced a nearly opaque solution with visible/near-infrared spectra containing transitions characteristic of fullerene anions. The identifying peak for C_{60} at 1,075 nm was readily observed in each sample. Filtration of the electro-reduced solutions through a 0.45- μ m membrane filter showed that almost all ($>98\%$) of the particles had dissolved, and laser desorption mass spectrometry of the filtered solutions confirmed that all varieties of empty and metallofullerenes, including both C_{74} and Gd@ C_{60} , had entered solution.

Upon reoxidation, the insoluble small-bandgap fullerenes and Gd-metallofullerenes polymerized on the working electrode and were separated from the remaining large-bandgap fullerenes which are soluble in benzonitrile as neutral species. Figure 2c shows a laser desorption mass spectrometry analysis of the electroplated film that formed on a Pt working electrode when the anion solution was reoxidized at +0.4 V. The purification of C_{74} , C_{80} , and higher insoluble fullerenes is dramatic; almost no detectable C_{60} remains in the sample. The electrochemical process is entirely reversible, and when the reoxidized film is placed in fresh benzonitrile, a purified anion solution identical to that in Fig. 2c is readily obtained after a second electrochemical reduction.

The anions could also be chemically reoxidized, and Fig. 3c shows the precipitate obtained upon addition of FcPF₆ (Fc, ferrocenium) to the electrochemically reduced Gd-metallofullerene anions. The Fc⁺ to Fc reduction occurs at about +0.1 V, and addition of FcPF₆ produced immediate oxidation and precipitation of the insoluble

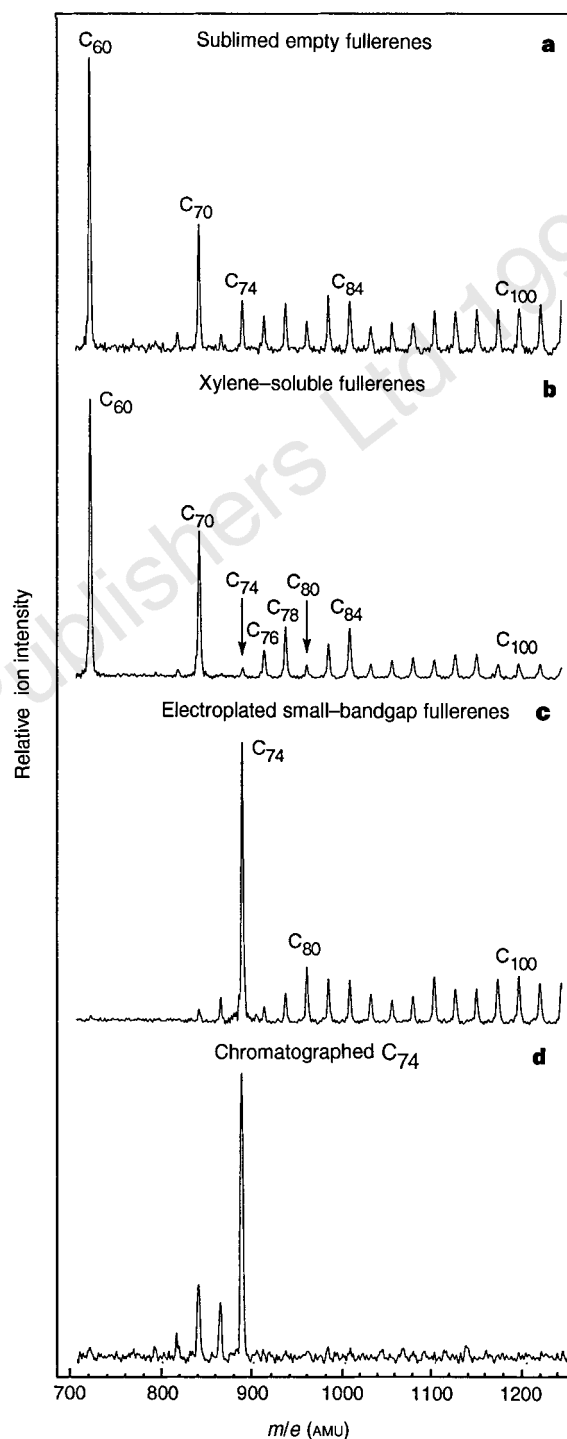


Figure 2 Positive ion laser desorption (Nd:YAG third harmonic at 355 nm) mass spectra of empty fullerenes at different process stages. Samples were placed onto target disks by either sublimation, solvent evaporation or electroplating. No desorption matrix was used. **a**, Fullerenes sublimed at 750 °C from soot. **b**, The xylene-soluble portion of the sublimate. The small amounts of C_{74} and C_{80} are probably due to fragmentation of higher fullerenes (by C_2 loss) during the laser desorption mass spectrometry (LDMS) process. **c**, Small-bandgap fullerene film electroplated on the working electrode by reoxidation of the solubilized anions. **d**, Chromatographic purification of C_{74} anions using a Regis 'Buckyclutcher' column (5 mm $\times$ 500 mm) with 6 ml min⁻¹ acetonitrile flow. After elution, the solution was dried, and the resulting salt was analysed by LDMS. Considerable fragmentation from the high laser fluence required to desorb and ionize the anions is evident.

small-bandgap fullerenes and Gd-metallofullerenes. With this type of reoxidation, a small amount of C_{60} is trapped in the precipitate.

Square-wave voltammetry is a powerful technique for probing the electronic structure of fullerenes, showing, for example, all six one-electron reductions that are expected from filling the t_{1u} LUMO of C_{60} (ref. 27). Square-wave voltammetry characterization of purified small-bandgap fullerenes and Gd-metallofullerenes (Fig. 4) reveals that their redox behaviour (and corresponding electronic structure) is substantially different from the soluble fullerenes¹⁰ and metallofullerenes²¹, as illustrated by the very small potential at which the anions return to the neutral state and polymerize. Oxidation to the neutral state at such small potentials ($V > -0.25$ V) indicates that the corresponding first reduction of the small-bandgap fullerenes occurs at a very low potential compared to soluble fullerenes, confirming their strong electronegativity and similarity to metallofullerenes. Cyclic voltammetry shows that

oxidations past +0.4 V are irreversible under these conditions, and suggests removal of electrons from orbitals below the HOMO.

The large HOMO–LUMO gap of C_{60} produces a large potential gap between the C_{60}^-/C_{60} and C_{60}/C_{60}^+ oxidations (the latter is irreversible under these conditions) (Fig. 4c). The correlation between the HOMO–LUMO gap and electrochemical potential gap has been studied for C_{60} , C_{70} , C_{76} , C_{78} and C_{84} , and the smallest observed potential difference was 1.47 V for $C_{78}(D_3)$ (ref. 10). Compared to these fullerenes, Fig. 4a confirms the small-bandgap assignment for C_{74} and the higher insoluble fullerenes.

Unlike open-shell fullerene anions such as C_{60}^{2-} or C_{70}^{2-} that precipitate immediately upon removal from the glovebox, the reduced small-bandgap fullerene and metallofullerene anion solutions appear to be stable in air for several hours, as verified by lack of precipitation and laser desorption mass spectrometry of the filtered solutions. A separation of C_{74} anions, using conventional

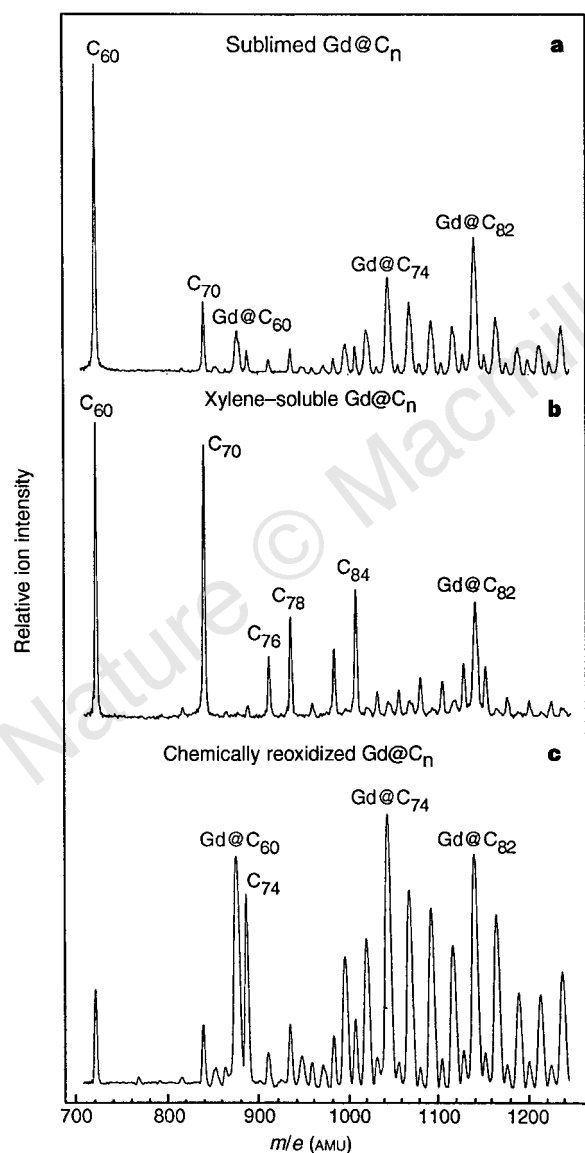


Figure 3 LDMS of Gd-metallofullerenes at different process stages. **a**, Gd-metallofullerenes sublimed at 750 °C from soot. Elemental analysis indicates that ~6–10% of the fullerenes contain Gd. **b**, The xylene-soluble portion of the sublimates. Only 4% of the Gd-metallofullerenes are soluble Gd@C₈₂. **c**, Precipitate obtained by chemical reoxidation of the solubilized anions with (ferricenium)PF₆.

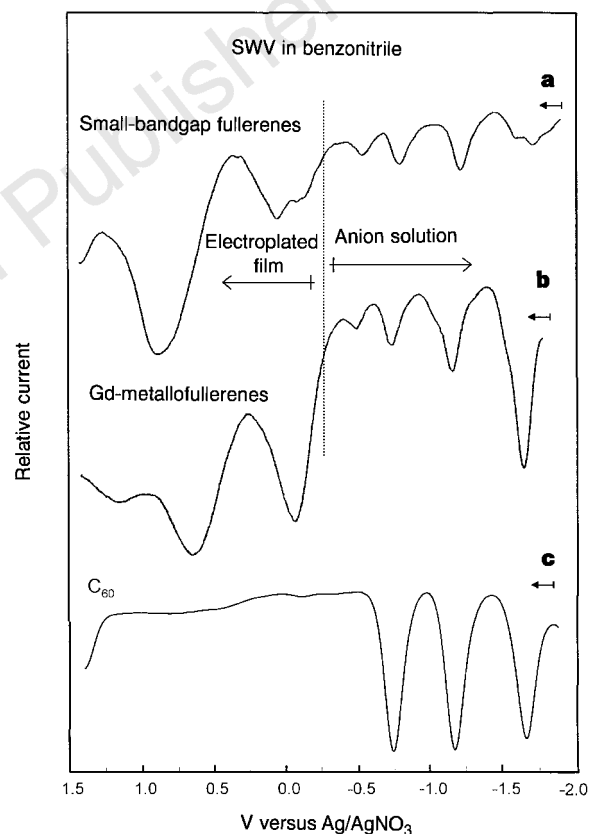


Figure 4 Square-wave voltammetry (SWV) of the different fullerene classes in 0.1 M $(n\text{-Bu})_4\text{NPF}_6/\text{BzCN}$ at room temperature. Owing to formation of polymeric films on the electrode, the SWV scans were started at -1.8 V and run positive to oxidize sequentially the solubilized anions. The dashed line shows the potential at which conversion of the anions into the polymerized film begins. The SWVs were generated using a frequency of 60 Hz, a pulse height of 50 mV, and a scan increment of 2 mV. **a**, Small-bandgap empty fullerenes with a composition similar to that shown in Fig. 2c. **b**, Gd-metallofullerenes with a composition similar to that shown in Fig. 3c. **c**, C_{60} , a soluble fullerene with a large 'electrochemical gap'. In contrast to the small-bandgap and Gd-metallofullerenes, the C_{60} remains in solution at all potentials until it is irreversibly oxidized at +1.45 V. Close inspection of the C_{60}^- oxidations shows that the oxidations of the metallofullerene and small-bandgap fullerene anions are not attributable to C_{60} . The similarity between sequential oxidation potentials among the different fullerenes shows that they conform closely to the charged-sphere model²⁰. This is particularly evident for the Gd-metallofullerene sample, which shows well defined SWV oxidation peaks even though the LDMS analysis (Fig. 3c) indicates that there are many distinct Gd-metallofullerene species.

high-performance liquid chromatography methods, is shown in Fig. 2d. This preliminary work suggests that, with further refinement, each of the small-bandgap and radical species, including $C_{80}(I_h)$ and $La@C_{60}$, can eventually be isolated and studied.

The electrochemical process can easily be scaled to gram quantities, and it should be possible to synthesize a wide range of new materials from small-bandgap fullerenes. As with conducting anti-ferromagnetic A_1C_{60} , polymerization precludes strong electron correlation to individual fullerenes, and the small HOMO–LUMO gap fullerenes should produce small or zero bandgap solids. It is likely that some of these, such as $C_{80}(I_h)$, will form three-dimensional organic conductors (or perhaps superconductors) without the need for external (or internal) dopants. Similarly, interesting electronic and magnetic properties are expected from the metallofullerene solids. □

Received 12 November 1997; accepted 30 March 1998.

1. Fowler, P. W. & Manolopoulos, D. E. *An Atlas of Fullerenes* (Clarendon, Oxford, 1995).
2. Dresselhaus, M. S., Dresselhaus, G. & Eklund, P. C. *Science of Fullerenes and Carbon Nanotubes* (Academic, San Diego, 1996).
3. Kratschmer, W., Lamb, L. D., Fostiropoulos, K. & Huffman, D. Solid C_{60} : a new form of carbon. *Nature* **347**, 354–356 (1990).
4. Taylor, R., Hare, J. P., Abdul-Sada, A. & Kroto, H. W. Isolation, separation and characterization of the fullerenes C_{60} and C_{70} : The third form of carbon. *J. Chem. Soc., Chem. Commun.* **20**, 1423–1425 (1990).
5. Diederich, F. & Whetten, R. L. Beyond C_{60} : The higher fullerenes. *Acc. Chem. Res.* **25**, 119–126 (1992).
6. Kikuchi, K. *et al.* Isolation and identification of fullerene family: C_{70} , C_{76} , C_{78} , C_{84} , C_{90} and C_{96} . *Chem. Phys. Lett.* **188**, 177–180 (1992).
7. Saito, S., Okada, S., Sawada, A. & Hamada, N. Common electronic structure and pentagon pairing in extractable fullerenes. *Phys. Rev. Lett.* **75**, 685–688 (1995).
8. Weaver, J. H. *et al.* Electronic structures of solid C_{60} : Experiment and theory. *Phys. Rev. Lett.* **66**, 1741–1744 (1991).
9. Hino, S., Kikuchi, K. & Achiba, Y. Photoelectron spectra of higher fullerenes and their potassium complexes. *Synth. Metals* **70**, 1337–1340 (1995).
10. Yang, Y. *et al.* Reversible fullerene electrochemistry: correlation with the HOMO–LUMO energy difference for C_{60} , C_{70} , C_{76} , C_{78} and C_{84} . *J. Am. Chem. Soc.* **117**, 7801–7804 (1995).
11. Liu, X., Schmalz, T. G. & Klein, D. J. Favorable structures for higher fullerenes. *Chem. Phys. Lett.* **188**, 550–554 (1992).
12. Zhang, B. L., Wang, C. Z., Ho, K. M. & Chan, C. T. The geometry of large fullerene cages: C_{72} to C_{100} . *J. Chem. Phys.* **98**, 3095–3102 (1993).
13. Xu, Z., Nakane, T. & Shinohara, H. Production and isolation of $Ca@C_{82}(I-IV)$ and $Ca@C_{84}(I-II)$ metallofullerenes. *J. Am. Chem. Soc.* **118**, 11309–11310 (1996).
14. Cummins, T. R. *et al.* Electronic states and molecular symmetry of the higher fullerene C_{80} . *Chem. Phys. Lett.* **261**, 228–233 (1996).
15. Nakao, K., Kurita, N. & Fujita, M. *Ab initio* molecular-orbital calculation for C_{70} and seven isomers of C_{60} . *Phys. Rev. B* **49**, 11415–11420 (1994).
16. Kobayashi, K., Nagase, S. & Akasaka, T. A theoretical study of C_{80} and $La_2@C_{80}$. *Chem. Phys. Lett.* **245**, 230–236 (1995).
17. Beer, F., Gugel, A., Martin, K., Rader, J. & Mullen, K. High-yield reactive extraction of giant fullerenes from soot. *J. Mater. Chem.* **7**, 1327–1330 (1997).
18. Chai, Y. *et al.* Fullerenes with metals inside. *J. Phys. Chem.* **95**, 7564–7568 (1991).
19. Wang, L. S. *et al.* The electronic structure of $Ca@C_{60}$. *Chem. Phys. Lett.* **207**, 354–359 (1993).
20. Boltalina, O. V., Ioffe, I. N., Sorokin, I. D. & Sidorov, L. N. Electron affinity of some endohedral lanthanide fullerenes. *J. Phys. Chem.* **101**, 9561–9563 (1997).
21. Suzuki, T. *et al.* Electrochemical properties of fullereneolanthanides. *Tetrahedron* **52**, 4973–4982 (1996).
22. Hummelen, J. C., Knight, B., Pavlovich, J., Gonzalez, R. & Wudl, F. Isolation of the heterofullerene $C_{59}N$ as its dimer $(C_{59}N)_2$. *Science* **269**, 1554–1556 (1995).
23. Gromov, A., Kratschmer, W. & Campbell, E. Extraction and HPLC purification of $Li@C_{60/70}$. *Chem. Commun.* **20**, 2003–2004 (1997).
24. Pekker, S. *et al.* Single-crystalline $(KC_{60})_n$: A conducting linear alkali fulleride polymer. *Science* **265**, 1077–1078 (1994).
25. Yeretizian, C. *et al.* Partial separation of fullerenes by gradient sublimation. *J. Phys. Chem.* **97**, 10097–10101 (1993).
26. Koh, W., Dubois, D., Kutner, W., Jones, M. T. & Kadish, K. M. Simultaneous cyclic voltammetry and electrochemical quartz crystal microbalance studies of buckminsterfullerene (C_{60}) film electrodeposition and tetra-*n*-butylammonium electrochromic doping in acetonitrile. *J. Phys. Chem.* **96**, 4163–4165 (1992).
27. Xie, Q., Perez-Cordero, E. & Echegoyen, L. Electrochemical Detection of C_{60}^{2-} and C_{70}^{2-} : Enhanced stability of fullerenes in solution. *J. Am. Chem. Soc.* **114**, 3978–3980 (1992).
28. Frisch, M. J. *et al.* Gaussian 94, Rev. E.1 (Gaussian, Inc., Pittsburgh, PA, 1995).
29. Green, W. H. Jr *et al.* Electronic structures and geometries of C_{60} anions via density functional calculations. *J. Phys. Chem.* **100**, 14892–14898 (1996).
30. Weaver, M. J. & Gao, X. Molecular capacitance: Sequential electron-transfer energetics for solution-phase metallic clusters in relation to gas-phase clusters and analogous interfaces. *J. Phys. Chem.* **97**, 332–338 (1993).
31. Haddon, R. C. Chemistry of the fullerenes: the manifestation of strain in a class of continuous aromatic molecules. *Science* **26**, 1545–1550 (1993).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements. We thank W. Bell, R. Cook, S. Dietz and J. Wright at TDA Research, L. Wilson at Rice University, and K. Veirs at LANL for productive technical discussions. This work was funded by a NSF SBIR grant to J.M.A.

Correspondence and requests for materials should be addressed to J.M.A.

An intermolecular $(H_2O)_{10}$ cluster in a solid-state supramolecular complex

Leonard J. Barbour, G. William Orr & Jerry L. Atwood

Department of Chemistry, University of Missouri-Columbia, Columbia, Missouri 65211, USA

Chemical self-assembly is the process by which 'programmed' molecular subunits spontaneously form complex supramolecular frameworks^{1,2}. This approach has been applied to many model systems, in which hydrogen bonds^{3,4}, metal–ligand coordination⁵ or other non-covalent interactions⁶ typically control the self-assembly process. In biology, self-assembly is generally dynamic and depends on the cooperation of many such non-covalent interactions. Water can play an important role in these biological self-assembly processes, for example by stabilizing the native conformation of biopolymers^{7–9}. Hydrogen-bonded $(H_2O)_n$ clusters^{10,11} can play an important role in stabilizing some supramolecular species, both natural and synthetic, in aqueous solution. Here we report the preparation and crystal structure of a self-assembled, three-dimensional supramolecular complex that is stabilized by an intricate array of non-covalent interactions involving contributions from solvent water clusters, most notably a water decamer $((H_2O)_{10})$ with an ice-like molecular arrangement. These findings show that the degree of structuring that can be imposed on water by its surroundings, and vice versa, can be profound.

We have been investigating supramolecular systems that contain molecular cavities as well as functional moieties present in biological macromolecules (such as amide groups and aromatic rings). An additional important consideration for such systems is that they have an affinity for water. To this end, we prepared the neutral, water-soluble, *exo*-bidentate ligand **1** from the reaction of 4-aminomethylpyridine with isophthaloyl chloride, and the coordination

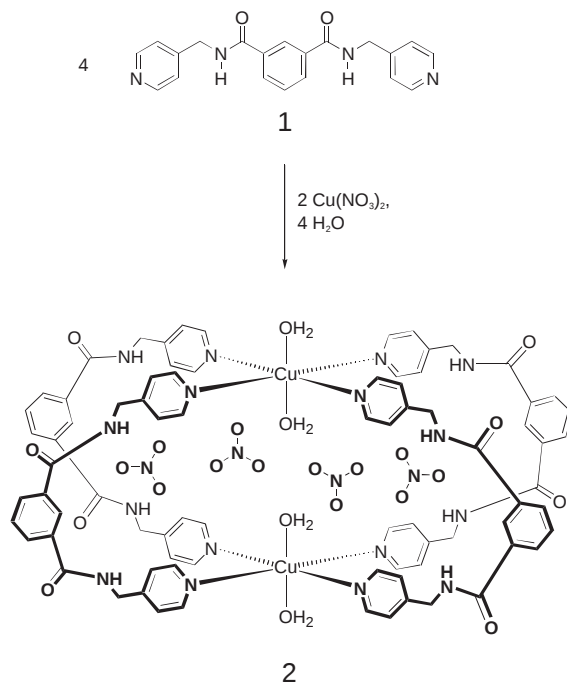


Figure 1 Schematic diagram representing the formation of a dinuclear cage-like supermolecule complex in aqueous media.

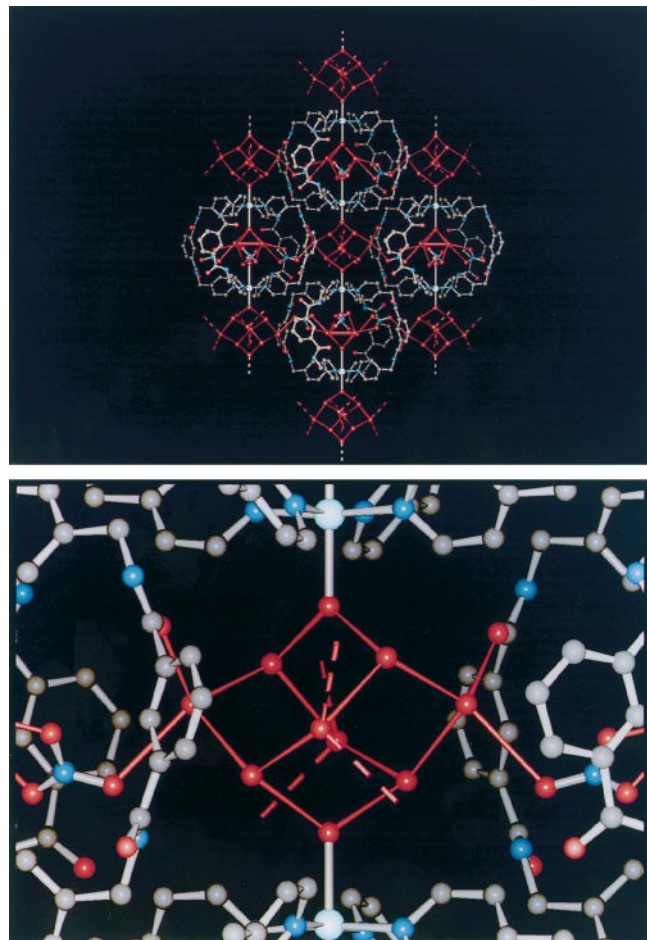


Figure 2 Ball-and-stick perspective view of sections of three adjacent supramolecular strands (viewed perpendicular to the crystallographic *c* axis). Colours are as follows: grey, carbon; red, oxygen; dark blue, nitrogen; light blue, copper. Hydrogen atoms are omitted for clarity, and hydrogen bonds are represented as solid red lines between oxygen atoms. Broken bonds indicate extension of the structure. As seen in the centre of the diagram (enlargement shown at bottom), the relative packing of the dinuclear tricyclic cage complexes results in the formation of the cavities which are occupied by the water decamers. The complex crystallizes in the tetragonal system, space group $I4_1/a$ (no. 88), $a = b = 21.9265(8) \text{ \AA}$, $c = 39.754(2) \text{ \AA}$, $V = 19112(1) \text{ \AA}^3$, $Z = 8$, $\rho_{\text{calc}} = 1.424 \text{ g cm}^{-3}$, $R_1 = 0.0962$ ($I > 2\sigma(I)$), $\chi(\text{Mo K}\alpha) = 0.70930 \text{ \AA}$. Details of the X-ray structure determination are available; see Supplementary Information.

complex **2** (as outlined in Fig. 1) by reacting **1** with $\text{Cu}(\text{NO}_3)_2$ in water.

The crystal structure of **2** (Fig. 2) reveals a dinuclear, macrotricyclic complex in which two octahedral Cu^{2+} ions are linked by four bridging ligands to form a molecular cage. All four equatorial positions on each metal centre are occupied by pyridyl groups while the axial positions are occupied by coordinated water molecules. In addition to two coordinated water molecules, the cage contains a hydrogen-bonded cluster consisting of at least six water molecules and four nitrate anions. The neutral supermolecules stack in linear arrays with their $\text{Cu}\cdots\text{Cu}$ axes parallel to $[001]$. Each array is surrounded by four identical nearest-neighbour arrays, and adjacent arrays are offset with respect to one another by half a structural unit. This staggered arrangement facilitates the formation of relatively large cavities between successive molecular cages within each linear array. Each of these cavities contains a discrete hydrogen-bonded water cluster consisting of ten water molecules (hydrogen bonds are inferred from short $\text{O}\cdots\text{O}$ contacts). Copper-coordinated water molecules constitute the extreme ends of the cluster with

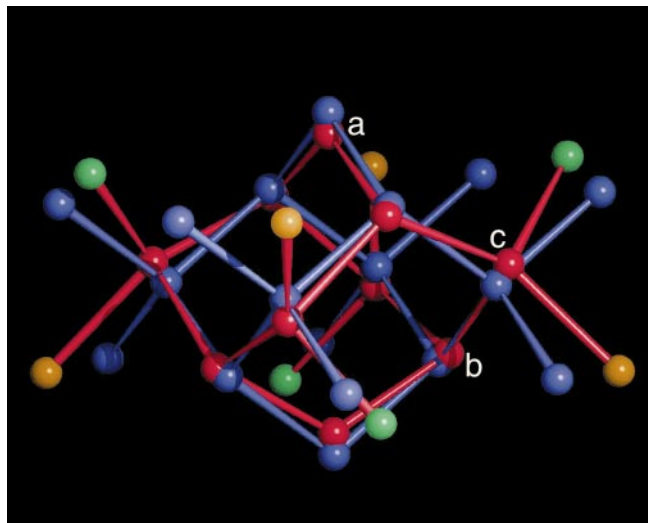


Figure 3 Comparison between the water decamer (in red) and ice I_c (in blue). The two structural units are similar in both their shape and dimensions (it should, of course, be noted that the water decamer is discrete while the ice structure is continuous). All bonds shown are hydrogen bonds. The water decamer is situated with its centre on a position of $\bar{4}$ site symmetry and the crystallographically unique water molecules are labelled a, b and c. Molecule a is coordinated to copper and forms a hydrogen bond to b which in turn hydrogen-bonds to c. Thus, a participates in the formation of one coordination bond and two hydrogen bonds, b participates in three hydrogen bonds and c participates in four hydrogen bonds. Of the latter, two hydrogen bonds are formed within the decamer and two with oxygen atoms of an amide group (green) and nitrate anion (orange) of an adjacent 'supermolecule'. The water decamer is slightly compressed along the $\text{Cu}\cdots\text{Cu}$ vector; the average $a\cdots a$ distance is $5.65 \text{ \AA}$ while the corresponding distance in ice I_c is $6.35 \text{ \AA}$. The greatest deviation from the ice structure is in the two unique $b\cdots b\cdots\text{O}_{\text{nitrate}}$ angles, which average 78° and 147° , respectively.

respect to the $[001]$ direction. Thus, in effect, the water decamer, as a supramolecular entity, fulfills the role of an *exo*-bidentate metal-bridging ligand.

Our initial aim had been to characterize the contents of the intramolecular cavity of the macrotricyclic coordination complex. Although this cleft contains an interesting arrangement of water molecules and nitrate anions, we shall limit the present discussion to the intermolecular water cluster which we believe to be of great significance to the current debate about the accretion of water in mixed-component systems. Figure 3 shows that the structure of the water decamer is remarkably similar to that of the smallest recognizable subunit of ice I_c (compare Fig. 4b). The structure contains two crystallographically unique water decamers located in similar environments. Each hydrogen bonded cluster is situated with its centroid on a position of $\bar{4}$ site symmetry and contains three unique internal bonds (average length, $2.80(5) \text{ \AA}$) and five unique internal angles (average angle, $110(9)^\circ$). In the structure¹² of ice I_c , determined at -130°C , the $\text{O}\cdots\text{O}$ distance is $2.75 \text{ \AA}$ and the $\text{O}\cdots\text{O}\cdots\text{O}$ angles are tetrahedral. The van der Waals volume available to the water decamer is $309 \text{ \AA}^3$, yielding an effective density of 0.967 g cm^{-3} for the cluster. Considering the obvious differences in environment, this value compares quite favourably with that of 0.934 g cm^{-3} obtained for ice I_c at -130°C and implies that the water cluster is very efficient at occupying the available space. Moreover, the eight hydrogen bonds formed between the water decamer and the amide groups and nitrate anions of adjacent linear arrays also contribute to the stabilization of the overall structure.

The literature relating to the structure of pure water in both the solid and liquid state, as well as in mixed-component systems, is extensive^{13,14}. Application of X-ray, neutron and electron diffraction

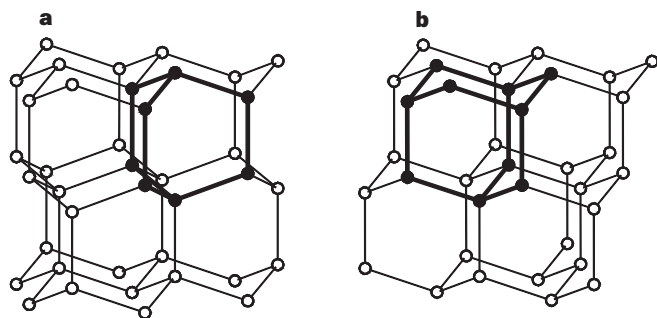


Figure 4 The structures of the ice phases I_h (a) and I_c (b). Oxygen atoms are shown as circles and hydrogen bonds are represented as lines. The darkened portions depict the smallest representative structural motif in each case. The hexagonal phase I_h is formed when water is cooled below its freezing point at atmospheric pressure. The cubic phase I_c can be formed by condensation of water vapour below -80°C , or via a phase transition from vitreous ice or ices II, III or V. Above -80°C , ice I_c itself transforms irreversibly to ice I_h . In both instances, each oxygen atom is situated at the centre of a tetrahedral arrangement of its four nearest-neighbour molecules. Its hydrogen atoms and lone pairs of electrons are directed towards these neighbours, thus facilitating the formation of four hydrogen bonds per water molecule, with each molecule acting as a double donor as well as a double acceptor. In each case, the lattice consists of two-dimensional layers of hexagonal rings in the chair conformation. These puckered sheets are stacked on one another, and are interconnected by the formation of hexagonal rings composed of three molecules each form adjacent layers. The two forms of ice I differ only in the stacking sequence of the puckered layers relative to one another, such that the interconnecting rings are in the boat conformation in ice I_h , while they assume the chair conformation in ice I_c .

techniques has afforded a relatively clear understanding of the structure of water in the solid state, and 12 polymorphic forms of ice, denoted as I_h , I_c and II–XI, have been characterized to date. Phases I_h and I_c are the two most common forms of ice and are strikingly similar in structure (Fig. 4). Extended water aggregates are also encountered in multi-component crystals and well-known examples include biological macromolecules¹⁵ and clathrate hydrates¹⁶. We also note that a recent study has produced a solid-state structure which is stabilized by an infinite two-dimensional ice-like framework intercalated between $\text{CdNi}(\text{CN})_4$ layers¹⁷.

An improved understanding of the three-dimensional structural aspects of water has important implications in the area of structural biology. Indeed, investigations of the forces at play in macromolecular interactions¹⁸ have demonstrated the importance of water structuring, and this point has been illustrated by several examples including the structures of *Scapharca* dimeric haemoglobin¹⁹, crambin²⁰, actinidin²¹ and carbonic anhydrase C (ref. 22). There is much evidence that alludes to the presence of ordered water clusters in the active clefts of these proteins, but unambiguous positional information is still rare. It is thought that water molecules contribute to the complex stability by mediating hydrogen bonds between the functional groups of the protein and the ligands, and by filling potential voids or holes inside the binding site²³.

The structure reported here demonstrates that the ice I_c arrangement can be a favourable conformation for a water cluster in a mixed-component system, even at room temperature. It also shows how the cluster and its surroundings assume a complementary relationship, resulting in the void available to the cluster being optimally occupied both in terms of packing efficiency and the maximization of intermolecular interactions. □

Received 5 February; accepted 6 April 1998.

- Lehn, J.-M. *Supramolecular Chemistry: Concepts and Perspectives* (VCH, Weinheim, 1995).
- Lehn, J.-M. *Supramolecular Chemistry*. *Science* **260**, 1762–1763 (1993).
- Seto, C. T. & Whitesides, G. M. Molecular self-assembly through hydrogen bonding: Supramolecular aggregates based on the cyanuric acid–melamine lattice. *J. Am. Chem. Soc.* **115**, 905–916 (1993).

- Wyler, R., de Mendoza, J. & Rebek, J. A synthetic cavity assembles through self-complementary hydrogen bonds. *Angew. Chem. Int. Edn Engl.* **32**, 1699–1701 (1993).
- Hasenknopf, B., Lehn, J.-M., Kneisel, B., Baum, O. G. & Fenske, D. Self-assembly of a circular double helicate. *Angew. Chem. Int. Edn Engl.* **35**, 1838–1840 (1996).
- Oriol, J. *et al.* A designed non-peptidic receptor that mimics the phosphocholine binding site of the McPC603 antibody. *Angew. Chem. Int. Edn Engl.* **35**, 1712–1715 (1996).
- Dehl, R. E. & Hoeve, C. A. Broad-line NMR study of H_2O and D_2O in collagen fibres. *J. Chem. Phys.* **50**, 3245–3251 (1969).
- Migchelsen, C., Berendsen, H. J. C. & Rupprecht, A. Hydration of DNA. Comparison of nuclear magnetic resonance results for oriented DNA in the A, B and C form. *J. Mol. Biol.* **37**, 235–237 (1968).
- Steckel, F. & Szapiro, S. Physical properties of heavy oxygen water. Part I—Density and thermal expansion. *Trans. Faraday Soc.* **59**, 331–343 (1963).
- Colson, S. D. & Dunning, T. H. The structure of Nature's solvent: water. *Science* **265**, 43–44 (1994).
- Liu, K., Cruzan, J. D. & Saykally, R. J. Water clusters. *Science* **271**, 929–933 (1996).
- König, H. A cubic ice modification. *Z. Kristallogr.* **105**, 279–286 (1944).
- Eisenberg, D. & Kauzmann, W. *The Structure and Properties of Water* (Oxford Univ. Press, New York, 1969).
- Franks, F. (ed.) *Water: A Comprehensive Treatise* Vols 1–4 (Plenum, New York, 1972).
- Westhoff, E. (ed.) *Water and Biological Macromolecules* (CRC Press, Boca Raton, 1993).
- Jeffrey, G. A. in *Inclusion Compounds* Vol. 1 (eds Atwood, J. L., Davies, J. E. D. & MacNicol, D. D.) Ch. 5 (Academic, London, 1984).
- Park, K.-M., Kurodo, R. & Iwamoto, T. A two-dimensional ice with the topography of edge-sharing hexagons intercalated between $\text{CdNi}(\text{CN})_4$ layers. *Angew. Chem. Int. Edn Engl.* **32**, 884–886 (1993).
- Rau, D. C. & Parsegian, V. A. Direct measurement of forces between linear polysaccharides Xanthan and Schizophyllan. *Science* **249**, 1278–1281 (1990).
- Royer, W. E., Pardanani, A., Gibson, Q. H., Peterson, E. S. & Feldman, J. M. Ordered water molecules as key allosteric mediators in a cooperative dimeric hemoglobin. *Proc. Natl Acad. Sci. USA* **93**, 14526–14531 (1996).
- Teeter, M. M. Water structure of a hydrophobic protein at atomic resolution: Pentagonal rings of water molecules in crystals of crambin. *Proc. Natl Acad. Sci. USA* **81**, 6014–6018 (1984).
- Baker, E. N. Structure of actinidin, after refinement at 1.7 Å resolution. *J. Mol. Biol.* **141**, 441–484 (1980).
- Liljas, A. *et al.* Crystal structure of human carbonic anhydrase C. *Nature New Biol.* **235**, 131–137 (1972).
- Quirocho, F. A., Wilson, D. K. & Vyas, N. K. Substrate specificity and affinity of a protein modulated by bound water molecules. *Nature* **340**, 404–407 (1989).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements. We thank C. Barnes for assistance with the unit cell and space group determination. This work was supported by the US NSF.

Correspondence and requests for materials should be addressed to J.L.A. (e-mail: chemja@mizzou1.missouri.edu).

Effect of the formation of the Isthmus of Panama on Atlantic Ocean thermohaline circulation

Gerald H. Haug* & Ralf Tiedemann

GEOMAR, Forschungszentrum für Marine Geowissenschaften, Universität Kiel, Wischhofstrasse 1-3, D-24148 Kiel, Germany

The Late Cenozoic closure of the seaway between the North and South American continents is thought to have caused extensive changes in ocean circulation and Northern Hemisphere climate^{1–2}. But the timing and consequences of the emergence of the Isthmus of Panama, which closed the seaway, remain controversial^{1–5}. Here we present stable-isotope and carbonate sand-fraction records from Caribbean sediments which, when compared to Atlantic and Pacific palaeoceanographic records, indicate that the closure caused a marked reorganization of ocean circulation starting 4.6 million years ago. Shallowing of the seaway intensified the Gulf Stream and introduced warm and saline water masses to high northern latitudes. These changes strengthened deep-water formation in the Labrador Sea over the next million years—as indicated by an increased deep-water ventilation and carbonate preservation in the Caribbean Sea—and favoured early Pliocene warming of the Northern Hemisphere. The evaporative cooling of surface waters during North Atlantic Deep Water formation would have introduced moisture to the Northern Hemisphere.

* Present address: Department of Earth Sciences, University of Southern California, Los Angeles, California 90089-0740, USA.

Although the pronounced intensification of Northern Hemisphere glaciation between 3.1 and 2.5 million years ago substantially lagged the full development of North Atlantic Deep Water formation, we propose that the increased atmospheric moisture content was a necessary precondition for ice-sheet growth, which was then triggered by the incremental changes in the Earth's orbital obliquity.

The gradual closing of the Isthmus of Panama lasted from 13 to 1.9 Myr ago^{3–5} (all originally published ages were adjusted to the new astronomically dated timescale⁶). Most evidence for restricted water-mass exchange through the Panama strait is based on sediment records from Caribbean and Pacific Deep Sea Drilling Program (DSDP) sites 502 and 503. Significant changes in planktonic foraminiferal assemblages occur at 6.8, 4.6, 2.5 and 1.9 Myr (ref. 4). A surface-water salinity increase in the Caribbean at 4.6 Myr is indicated by $\delta^{18}\text{O}$ values of planktonic foraminifera¹ and implies a shoaling of the seaway to <100 m water depth. Shallow-water fossils from both sides of the Panama–Costa Rica region indicate⁷ that the closure was almost complete at 3.6 Myr, but the final closure allowing land mammal exchange was achieved⁸ at 2.7 Myr, coincident with the glacial-induced sea-level drop during the main

intensification of Northern Hemisphere ice-sheet growth. However, the identification of the particular step in the closure of the Panamanian gateway that acted as a critical threshold for profound changes in deep-ocean circulation and climate remained qualitative and speculative⁹. Here we present proxy data which demonstrate that the closure has affected deep ocean circulation since 4.6 Myr ago.

Today, a mixture of nutrient-enriched, low- $\delta^{13}\text{C}$ Antarctic Intermediate Water (AAIW) and nutrient-depleted, high- $\delta^{13}\text{C}$ Upper North Atlantic Deep Water (UNADW) and Mediterranean Overflow Water cross the Atlantic–Caribbean sills at 1,600–1,900 m (Windward Passage, Anegada–Jungfern Passage) and fill the deep Caribbean basins. During the past 2.5 Myr, the relative proportion of northern- and southern-component water masses were related to glacial–interglacial differences in the formation rate of UNADW¹⁰. A weaker UNADW formation during interglacials led AAIW to extend further north and result in a less ventilated, more corrosive Caribbean deep water. Hence, the Caribbean Sea is a highly sensitive recorder of ventilation changes in the upper Atlantic if the sill depth remained constant. Tectonic evidence from the Lesser Antilles arc and Aves ridge suggests no significant vertical movements since the middle Miocene (20–15 Myr ago) when a thick crust was established; vertical movements of <100 m Myr^{–1} are expected, which were likely to have been closer to a few metres per Myr (ref. 11).

We report epibenthic foraminiferal $\delta^{18}\text{O}$, $\delta^{13}\text{C}$ and percentage sand records of the carbonate fraction from ODP Site 999 (12° 44' N, 78° 44' W, Colombian basin, water depth 2,828 m) for the time interval 2.0–5.3 Myr. The $\delta^{13}\text{C}$ values of *Cibicidoides wuellerstorfi* are a proxy for deep-water ventilation¹², as $\delta^{13}\text{C}$ of sea water is closely linked to seawater nutrient and oxygen levels, with higher $\delta^{13}\text{C}$ values indicating lower nutrient concentrations and better ventilation¹³. The sand content (>63 μm) of deep-sea carbonates is a sensitive indicator of changes in carbonate dissolution. The sand content (foraminifera shells) decreases as dissolution progresses¹⁴. The $\delta^{18}\text{O}$ of *C. wuellerstorfi* is a proxy for changes in continental ice volume and deep water temperature. The age model of Site 999 is based on $\delta^{18}\text{O}$ stratigraphy, and was correlated to the astronomically dated $\delta^{18}\text{O}$ records from equatorial east Pacific Site 846 (ref. 6) and equatorial east Atlantic Site 659 (ref. 15).

Oceanographic conditions that result in changes in both $\delta^{13}\text{C}$ and sand contents are documented in Figs 1 and 2. Before 4.6 Myr, low epibenthic $\delta^{13}\text{C}$ values and low sand contents indicate a poorly ventilated deep Caribbean and severe carbonate dissolution. In the early Pliocene, similar low $\delta^{13}\text{C}$ values of ~0.2‰ have been

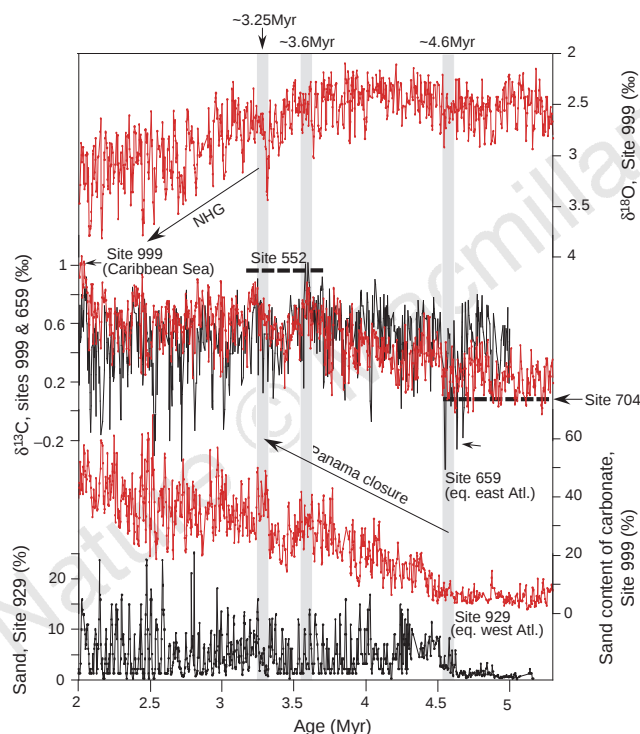


Figure 1 Palaeoceanographic records spanning the time interval 2–5.3 Myr. Top, the Site 999 isotope record (red), derived from analysing three specimens of the epibenthic foraminifera *Cibicidoides wuellerstorfi* (or if not available, *Cibicidoides kullenbergi*) of the size fraction 315–500 μm on a Finnigan MAT 252 in the stable-isotope laboratory at Geomar Kiel. $\delta^{18}\text{O}$ values are adjusted to sea water equilibrium by +0.64‰. The timescale for ODP Site 999 was achieved by adjusting the $\delta^{18}\text{O}$ fluctuations to the astronomical calibrated benthic $\delta^{18}\text{O}$ records of sites 846 (ref. 6) and 659 (ref. 15). Middle, carbon isotope records of sites 999 (red) and 659 (grey; equatorial east Atlantic; 18° 05' N, 21° 02' W; water depth 3,070 m). Data of both records are based on the epibenthic foraminifera *C. wuellerstorfi*. The dotted lines illustrate the average $\delta^{13}\text{C}$ values of sites 552 (ref. 17) (North Atlantic, water depth 2,301 m) and 704 (ref. 16) (Southern Ocean, water depth 2,532 m). Bottom, the sand contents of sites 999 (red) and 929 (grey) (coarse fraction > 63 μm); data from refs 14, 19, 33. At Site 999, the sand content is given as the percentage of total carbonate (determined with a LECO-Analyzer). The arrows mark the Pliocene cooling trend between 3.1 and 2.5 Myr (intensification of Northern Hemisphere glaciation) and the main step in Panamanian closure.

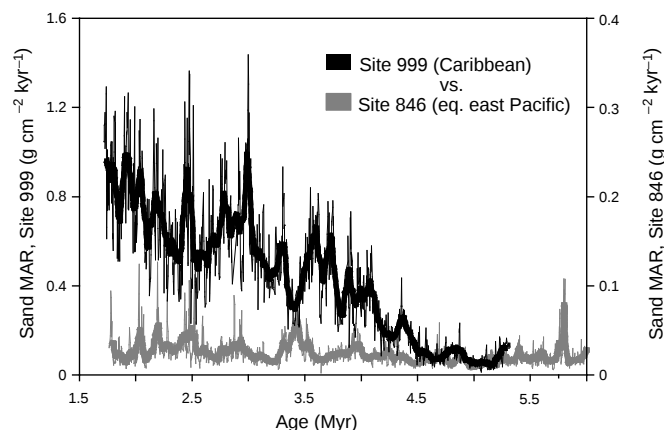


Figure 2 Comparison of carbonate sand-fraction mass accumulation rates (sand MAR) at sites 999 (left axis) and 846 (ref. 6) (right axis) plotted versus age for the interval 2–5.3 Myr. MAR are calculated by using GRAPE density values³⁴. Thin lines are original data, and thick lines are data smoothed by a 20-point running mean.

documented only at subantarctic South Atlantic Site 704 (ref. 16) (2,532 m water depth), in contrast to higher North Atlantic values of $\sim 1\text{‰}$ (for example, sites 659, 552 (ref. 17)). This suggests that the Caribbean deep water was dominated by a $\delta^{13}\text{C}$ -depleted Southern Ocean water mass (AAIW) before 4.6 Myr. After 4.6 Myr, deep-water ventilation as well as carbonate preservation increased into the late Pliocene due to a deepening of the lysocline. This is interpreted to reflect a progressively stronger influence of less corrosive and $\delta^{13}\text{C}$ -enriched northern component water due to an increase in UNADW formation. This increase at 4.6 Myr is paralleled by an increased formation of Lower North Atlantic Deep Water (LNADW) as indicated by records of deep-water ventilation (Fig. 1) and carbonate preservation in the equatorial east (ODP sites 659 and 665 (ref. 18)) and west Atlantic (Ceara rise depth transect, sites 925–929 (ref. 19)) below 3,000 m water depth.

A first ventilation maximum in the Caribbean Sea as well as in the deep Atlantic was reached at 3.6 Myr, when Caribbean $\delta^{13}\text{C}$ values approached those from North Atlantic component water (Site 659, Fig. 1). This mechanism supplied additional heat and moisture to the Northern Hemisphere and may have contributed to the mid-Pliocene warmth. Since 3.6 Myr, both sites 999 and 659 show similar $\delta^{13}\text{C}$ maxima and reflect the increased strength of northern component water masses. During cooler periods, $\delta^{13}\text{C}$ minima at Site 659 are more pronounced than those from Caribbean Site 999 and reflect vertical fluctuations of the LNADW-AABW mixing zone (AABW, Antarctic Bottom Water). Thus, even though LNADW may have been in the 'reduced' mode during harsh climate episodes, the Caribbean was still relatively nutrient-poor compared to the deep Atlantic (Site 659) because of increased formation of UNADW. This suggests that the familiar dipole in the Pleistocene ocean circulation^{10,20} has been operating at least since 3.6 Myr ago.

The comparison between the Caribbean and Pacific sand-fraction records of sites 999 and 846 demonstrates a substantial change in ocean carbonate preservation at 4.6 Myr (Fig. 2). Before 4.6 Myr, vigorous carbonate dissolution characterized both the Caribbean and the Pacific. Beginning at 4.6 Myr, the increasing thermohaline circulation amplified the inter-basin fractionation between the Atlantic and Pacific, which is reflected in a strong increase in Caribbean and Atlantic¹⁹ carbonate preservation and a remaining strong carbonate dissolution in the Pacific. This sedimentary response to changes in ocean circulation before and after isthmus formation has been predicted by ocean model studies²¹.

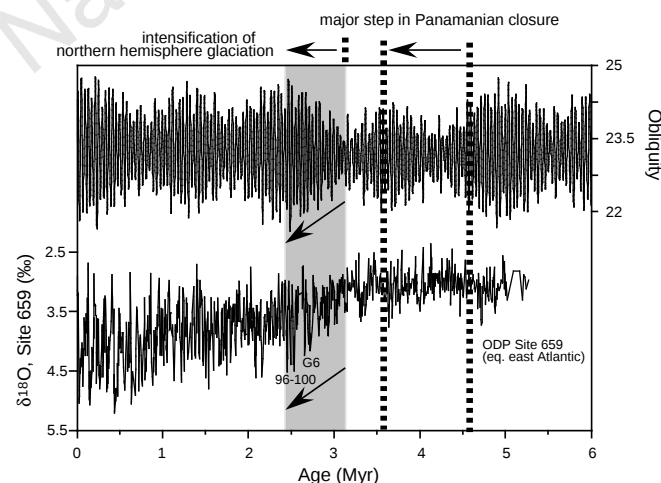


Figure 3 Calculated obliquity amplitude fluctuations³¹ and the benthic oxygen isotope record of Site 659¹⁵ versus age for the past 6 Myr. The arrows mark the Pliocene cooling trend between 3.1 and 2.5 Myr which resulted in the first large ice-sheet build-up in the Northern Hemisphere during isotopic stages G6-96 (labelled).

Our data provide a precise picture of the final phase of NADW intensification that had been developing during the mid-Miocene²². In response to the gradual emergence of the Central American seaway, we observe an enhanced thermohaline overturn since 4.6 Myr that reached a first maximum at 3.6 Myr. This was amplified by an increased salt transport to the North Atlantic and the initiation or intensification of UNADW formation in the Labrador Sea, as predicted by recent results²³ from global circulation model simulations. In support of this interpretation, we note that results from the Labrador Sea (ODP Leg 105) indicate increased bottom-water currents and drift sedimentation since ~ 4.5 Myr (ref. 24).

The closure of the Panamanian seaway has always been an attractive candidate for the ultimate cause of the Pliocene intensification of the Northern Hemisphere glaciation⁹. The pronounced ice-sheet growth in Eurasia, Greenland and North America is marked by a progressive ^{18}O -enrichment in benthic foraminifera $\delta^{18}\text{O}$ records between 3.1 and 2.5 Myr (Fig. 3) and by the massive appearance of ice-rafted debris in northern high-latitude oceans since 2.7 Myr (ref. 25). The intensification of Northern Hemisphere glaciation finalizes the Cenozoic cooling trend, which started in the early Eocene and is marked by first indications of ice sheets in Antarctica 36 Myr ago²⁶. This long-term cooling is considered to be a direct response to permanent removal of atmospheric CO_2 through enhanced silicate weathering²⁷ and/or enhanced burial of organic carbon²⁸ resulting from tectonically uplifted areas such as the Himalayas and American West. This long-term cooling brought the climate system of the Earth to a state critical for ice-sheet build-up in the Northern Hemisphere. This has been the case since ~ 10 –5 Myr ago, when the first, and minor, occurrence of ice-rafted debris in the Arctic and North Atlantic indicates the first attempts of the climate system to start a glaciation²⁹. However, until 2.7 Myr ago, the climate system failed to amplify and continue a large Northern Hemisphere glaciation.

To initiate and continue the build-up of the prominent Laurentide and Scandinavian ice sheets, three factors are needed to act together. First, general cooling must have reached a critical threshold to allow precipitation to fall as snow rather than rain. Second, moisture needs to be introduced to high northern latitudes. Our results suggest that moisture was provided by an increased thermohaline circulation and Gulf Stream flow since 4.6 Myr, well before the intensification of Northern Hemisphere glaciation. Third, astronomical theory requires that the summer in northern high latitudes must be cold enough to prevent winter snow from melting³⁰. High-amplitude fluctuations in the Earth's obliquity (low tilt angle) triggered cold summers in the Northern Hemisphere, and prepared the way for strengthening of the glacial-interglacial 41-kyr cycles during late Pliocene and early Pleistocene^{15,30}. However, a pronounced long-term minimum in obliquity amplitude fluctuations occurred between 4.5 and 3.1 Myr (ref. 31). The $\delta^{18}\text{O}$ records of sites 659 (ref. 15), 846 (ref. 6) and 999 show that during this unfavourable orbital configuration there may have been several failed attempts of the climate system to start the glaciation, for example during 4.1–3.9 Myr and 3.5–3.3 Myr. We therefore suggest that the progressive increase in obliquity amplitudes between 3.1 and 2.5 Myr was the final trigger for amplification and continuation of the long-term expansion of Northern Hemisphere ice sheets after the necessary preconditions were met 4.6–3.6 Myr ago by formation of the Isthmus of Panama. These incremental changes in obliquity³², coupled with changes in the background state of the ocean, suggest a threshold value for ice-sheet growth, which should be testable with climate models. □

Received 28 July 1997; accepted 14 April 1998.

1. Keigwin, L. D. Isotope paleoceanography of the Caribbean and east Pacific: role of Panama uplift in late Neogene time. *Science* **217**, 350–353 (1982).
2. Maier-Reimer, E., Mikolajewicz, U. & Crowley, T. J. Ocean general circulation model sensitivity experiment with an open American Isthmus. *Paleoceanography* **5**, 349–366 (1990).
3. Duque-Caro, H. Neogene stratigraphy, paleoceanography and paleobiology in northwest South America and the evolution of the Panama Seaway. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **77**, 203–234 (1990).

4. Keller, G., Zenker, C. E. & Stone, S. M. Late Neogene history of the Pacific–Caribbean gateway. *J. S. Am. Earth Sci.* **2**, 73–108 (1989).
5. Collins, L. S., Coates, A. G., Berggren, W. A., Aubry, M.-P. & Zhang, J. The late Miocene Panama isthmian strait. *Geology* **24**, 687–690 (1996).
6. Shackleton, N. J., Hall, M. & Pate, D. Pliocene stable isotope stratigraphy of Site 846. *Proc. ODP Sci. Res.* **138**, 337–357 (1995).
7. Coates, A. G. et al. Closure of the Isthmus of Panama: The near-shore marine record of Costa Rica and western Panama. *Geol. Soc. Am. Bull.* **104**, 814–828 (1992).
8. Marshall, L. G. Land mammals and the Great American Interchange. *Am. Sci.* **76**, 380–388 (1988).
9. Hay, W. W. Tectonics and climate. *Geol. Rundsch.* **85**, 409–437 (1996).
10. Oppo, D. W. et al. A $\delta^{13}\text{C}$ record of Upper North Atlantic Deep Water during the past 2.6 million years. *Paleoceanography* **10**, 372–394 (1995).
11. Driscoll, N. W. & Diebold, J. B. Tectonic and stratigraphic development of the eastern Caribbean: New constraints from multichannel seismic data. In *Caribbean Sedimentary Basins* (ed. Mann, P.) (Elsevier Spec. Publ., in the press).
12. McCorkle, D. C. & Keigwin, L. D. Depth profiles of $\delta^{13}\text{C}$ in bottom water and core top *C. wuellerstorfi* on the Ontong Java Plateau and Emperor Seamounts. *Paleoceanography* **9**, 197–208 (1994).
13. Kroppnick, P. M. The distribution of ^{13}C of ΣCO_2 in the world ocean. *Deep-Sea Res.* **32**, 57–84 (1985).
14. Bickert, T., Cordes, R. & Wefer, G. Late Pliocene to mid-Pleistocene (2.6–1.0 Ma) carbonate dissolution in the western equatorial Atlantic: Results of ODP Leg 154, Ceara Rise. *Proc. ODP Sci. Res.* **154** (1994).
15. Tiedemann, R., Sarntheim, M. & Shackleton, N. J. Astronomic timescale for the Pliocene Atlantic $\delta^{18}\text{O}$ and dust flux records of ODP Site 659. *Paleoceanography* **9**, 619–638 (1994).
16. Hodell, D. A., Müller, D. W., Ciesielski, P. F. & Mead, G. A. Synthesis of oxygen and carbon isotopic results from Site 704: Implications for major climatic-geochemical transitions during the late Neogene. *Proc. ODP Sci. Res.* **114**, 475–480 (1991).
17. Shackleton, N. J. & Hall, M. A. Oxygen and carbon isotope stratigraphy of DSDP Hole 552A: Pliocene–Pleistocene glacial history. *Init. Rep. DSDP* **81**, 599–610 (1984).
18. Ruddiman, W. F. et al. *Proc. ODP Init. Rep.* **108** (1988).
19. Tiedemann, R. & Franz, S. O. Deep-water circulation, chemistry, and terrigenous sediment supply in the equatorial Atlantic during the Pliocene, 3.3–2.6 Ma and 5–4.5 Ma. *Proc. ODP Sci. Res.* **154**, 299–318 (1997).
20. Boyle, E. A. & Keigwin, L. D. North Atlantic thermohaline circulation during the past 20,000 years linked to high-latitude surface temperature. *Nature* **330**, 35–40 (1987).
21. Heinze, C. & Crowley, T. J. Sedimentary response to ocean gateway circulation changes. *Paleoceanography* **12**, 742–754 (1997).
22. Wright, J. D. & Miller, K. G. Control of North Atlantic Deep Water circulation by the Greenland–Scotland Ridge. *Paleoceanography* **11**, 157–171 (1996).
23. Mikolajewicz, U. & Crowley, T. J. Response of a coupled ocean/energy balance model to restricted flow through the central American Isthmus. *Paleoceanography* **12**, 429–441 (1997).
24. Arthur, M. A. et al. Seismic stratigraphy and history of deep circulation and sediment drift development in Baffin Bay and the Labrador Sea. *Proc. ODP Sci. Res.* **105**, 957–989 (1989).
25. Shackleton, N. J. et al. Oxygen isotope calibration of the onset of ice-raftering and history of glaciation in the North Atlantic region. *Nature* **307**, 620–623 (1984).
26. Miller, K. G., Fairbanks, R. G. & Mountain, G. S. Tertiary oxygen isotope synthesis, sea level history, and continental margin erosion. *Paleoceanography* **2**, 1–19 (1987).
27. Raymo, M. E. & Ruddiman, W. F. Tectonic forcing of late Cenozoic climate. *Nature* **359**, 117–122 (1992).
28. France-Lanord, C. & Derry, L. A. Organic carbon burial forcing of the carbon cycle from Himalayan erosion. *Nature* **390**, 65–67 (1997).
29. Myhre, A. M. et al. *Proc. ODP Init. Rep.* **151**, 1–926 (1996).
30. Imbrie, J. et al. On the structure and origin of major glaciation cycles, 1. linear responses to Milankovitch forcing. *Paleoceanography* **7**, 701–738 (1992).
31. Laskar, J., Joutel, F. & Boudin, F. Orbital, precessional, and insolation quantities for the Earth from –20 Myr to +10 Myr. *Astron. Astrophys.* **270**, 522–533 (1993).
32. Li, X. S. et al. Simulating late Pliocene Northern Hemisphere climate with the LLN 2-D model. *Geophys. Res. Lett.* (in the press).
33. Billups, K., Ravelo, A. C. & Zachos, J. C. Early Pliocene deep-water circulation: Stable isotope evidence for enhanced northern component deep water. *Proc. ODP Sci. Res.* **154**, 319–330 (1997).
34. Sigurdsson, H. et al. *Proc. ODP Init. Rep.* **165**, 1–674 (1997).

Acknowledgements. We thank R. Zahn, S. Franz, L. Keigwin, D. Oppo, N. Driscoll, R. Norris, J. McManus and T. Bickert for discussions and criticism, and S. Hustedt and A. Richter for technical assistance. This work was supported by the Deutsche Forschungsgemeinschaft (DFG).

Correspondence and requests for materials should be addressed to G.H.H. (e-mail: ghaug@earth.usc.edu).

Palaeozoic and Proterozoic zircons from the Mid-Atlantic Ridge

Joachim Pilot, Carl-Dietrich Werner, Frank Haubrich & Nils Baumann

Institute of Mineralogy, Freiberg University of Mining and Technology, D-09596 Freiberg, Germany

According to the theory of plate tectonics, rocks found in the vicinity of mid-ocean ridges—where oceanic plates are created—should be relatively young (at most several Myr old). Here we report the discovery of zircons with ages of about 330 and 1,600 Myr that were drilled from exposed gabbros beneath the Mid-Atlantic Ridge near the Kane fracture zone^{1–4}. Age determi-

nations were made using the $^{207}\text{Pb}/^{206}\text{Pb}$ evaporation method⁵ and confirmed with conventional U–Pb dating and ion microprobe (SHRIMP) analysis. We suggest two plausible explanations for the origin of these unusually old zircons. During the opening of the Atlantic, sheared crustal material or delaminated continental lithosphere sank into small roll-like circulation cells^{6,7} that developed in the shallow mantle at each side of the ridge axis and the material was then transported through these cells to the ridge axis. Alternatively, material from the continental crust has been trapped within the Kane fracture zone since the opening of the Atlantic Ocean basin through a series of transform migrations and ridge jumps^{8,9}, with portions of this material subsequently migrating down the ridge axis.

Leg 153 of the Ocean Drilling Program drilled five sites on the western valley wall of the Mid-Atlantic Ridge (MAR) between 23° 20' N and 23° 30' N, 5–35 km south of the Kane transform (Fig. 1). This area is located in the central Atlantic Ocean, ~2,000 km from the continental margins and far from any islands. Leg 153 recovered harzburgites of the upper mantle which are depleted by partial melting followed by the formation of gabbroic melts^{1–4}. The peridotite–harzburgite–gabbro complex is interpreted to be uplifted and exhumed by normal faults^{1–4,10,11}. The gabbros comprise normal gabbro, olivine gabbro, gabbro-norite, olivine gabbro-norite, olivine norite, and troctolite.

From the recovered cores, we selected 30 gabbro samples from material drilled 7.7–68.9 m below the sea floor. These samples were the least altered ones and did not contain visible schlieren and veins.

While performing oxygen-isotope investigations of these gabbro samples, we discovered numerous zircons. Some of these samples contain 1–4 zircons in 5 cm³; however, close to 40 zircons were found in sample 137. The number of zircons identified does not correspond to the Zr content of the host rock (mean ~20 p.p.m.). The morphology of zircons in samples from the same borehole is similar, but differs in samples from different boreholes. Most of the zircons are 20–60 μm in size, yellow-white and transparent with some reddish tear-type inclusions, and sometimes show a lineation parallel to crystal faces (Fig. 2). Frequent types in the Pupin diagram¹² are P1, P2, G1 and G2, with short prismatic shapes being most abundant. Sample 122 from borehole 922A contained zircons that were of irregular, or had no discernible, shape.

Zircons in gabbro complexes which occur mainly in veinlets and schlieren have been reported several times^{10,11,13–15}. In thin sections such zircons are seen at the borders between mineral grains; in one case, a zircon was included in an ilmenite grain¹³. However, our sampling avoided visible veins, schlieren and any signs of alteration as far as possible; the zircons used for our age determinations are clearly different from those which may be formed in the highly differentiated rest-magma of the gabbroic melts.

If the zircons were formed during the generation of the gabbroic magma, then their age would be expected to be ~1 Myr (ref. 1); because of their U contents it would be difficult to determine a U–Pb age. To reveal the history of these zircons, we first made $^{207}\text{Pb}/^{206}\text{Pb}$ age determinations of carefully selected single zircons from three samples (samples 134 and 137, borehole 923A; sample 122, borehole 922A), using the well established mass-spectrometric evaporation method⁵.

All zircons contain both radiogenic lead and small, but significant,

Table 1 Isotope ratios of zircons from sample 122

Sample	$^{204}\text{Pb}/^{206}\text{Pb}_m$	$^{207}\text{Pb}/^{206}\text{Pb}_m$	$^{208}\text{Pb}/^{206}\text{Pb}_m$	$^{206}\text{Pb}/^{238}\text{U}$	$^{207}\text{Pb}/^{235}\text{U}$
122/L14-1	0.002401(18)	0.13038(17)	0.14305(23)	0.2022(16)	2.714(37)
122/L14-2	0.002606(16)	0.12967(12)	0.13313(27)	0.1821(25)	2.353(45)
122/L14-3	0.002516(32)	0.13605(12)	0.20123(24)	0.2552(21)	3.584(58)

Shown are lead isotope ratios and corrected $^{206}\text{Pb}/^{238}\text{U}$ and $^{207}\text{Pb}/^{235}\text{U}$ ratios from sample 122 zircons (second separation). (Subscript m indicates measured ratio but mass spectrometric discrimination is corrected).

amounts of common lead. During the first step in this analytical procedure, weakly bound lead is evaporated from the zircons. Because in the second step the Pb in the zircons is very radiogenic, we plotted the measured $^{207}\text{Pb}/^{206}\text{Pb}$ ratios against the measured $^{204}\text{Pb}/^{206}\text{Pb}$ ratios¹⁶ (Fig. 3). We obtained two trends with straight regression lines. The resulting radiogenic $^{207}\text{Pb}^*/^{206}\text{Pb}^*$ ratio of the lower line corresponds to an age of 330 ± 17 Myr, and the ratio for

the upper line corresponds to an age of $1,623 \pm 12$ Myr (errors show 95% confidence level).

We note that these two radiogenic $^{207}\text{Pb}^*/^{206}\text{Pb}^*$ ratios could also develop in young zircons of around 0.5 and 0.15 Myr owing to radioactive non-equilibrium in the U-chains^{17–19}. But an estimation of the Pb content of some of the more Pb-rich zircons using the ^{206}Pb currents showed that these zircons must be far older than

Figure 1 Generalized bathymetric and tectonic map of the Mid-Atlantic Ridge south of the Kane transform. Included are the location of Ocean Drilling Program Leg 153 sites 921, 922 and 923 at the western valley wall. The dark grey areas are uplifted units of the periodotite-harzburgite-gabbro complex.

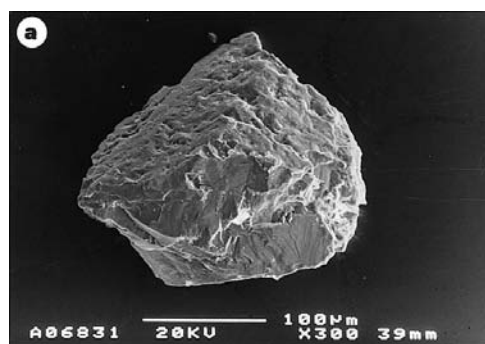
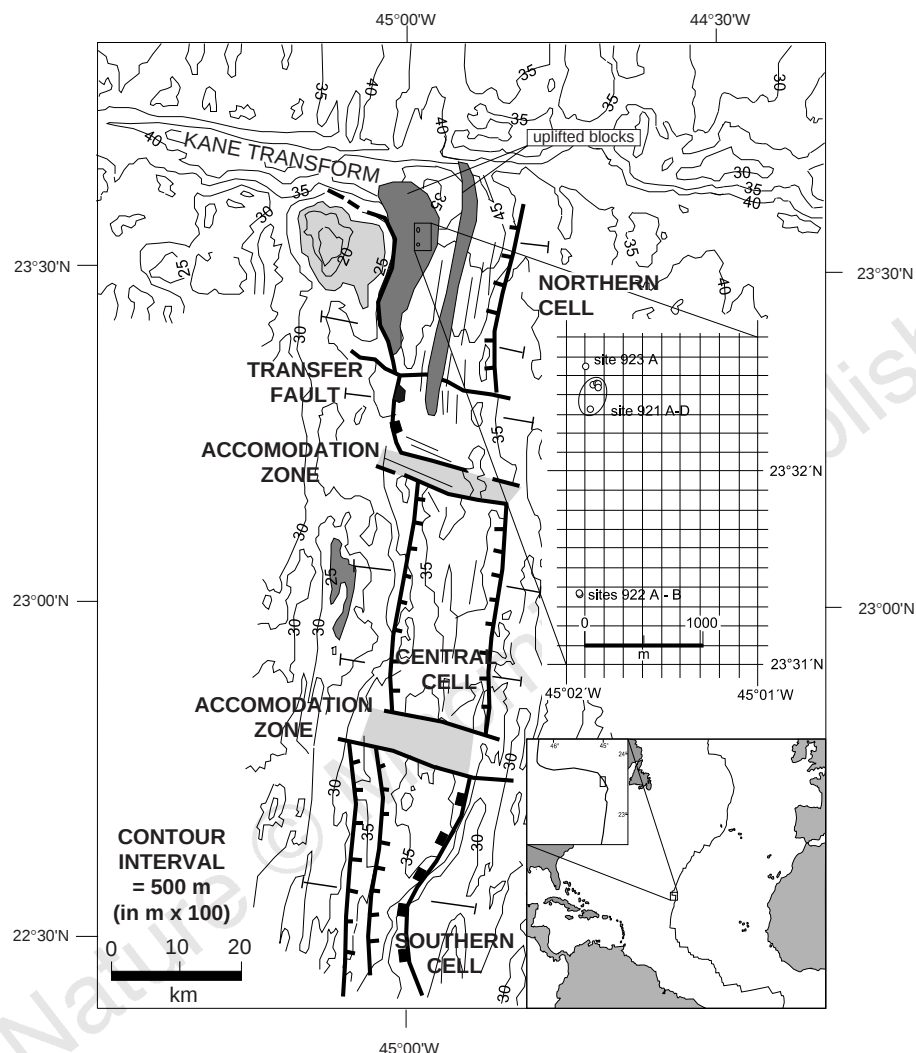


Figure 2 Scanning electron micrographs of two selected zircons. **a**, Zircon with irregular shape (hole 922A, sample 122); **b**, zircon with middle outer part broken off, so that the internal texture is visible (hole 923A, sample 134). This texture may be explained by strong tectonic stress, internal reorganization and temperature change. In the recovery of all zircons, great care was taken during breaking and crushing the core samples (~50 g), and during sieving of grain size fractions, to avoid any possibility of cross-contamination. All equipment was carefully cleaned before processing a new sample, and all 30 samples were processed in a

continuous sequence (the three samples used for age determinations were at places 18, 27 and 29). Heavy-mineral concentrations were obtained using bromoform and magnetic separation. The zircons were hand-picked under a microscope. From sample 122 we repeated the heavy-mineral separations; zircons of both separations were dated and coincide in age. Sample 134 is from the same borehole as the 10-m-deeper sample 137; both gave the same age. Scale bars: **a**, 100 μm ; **b**, 10 μm .

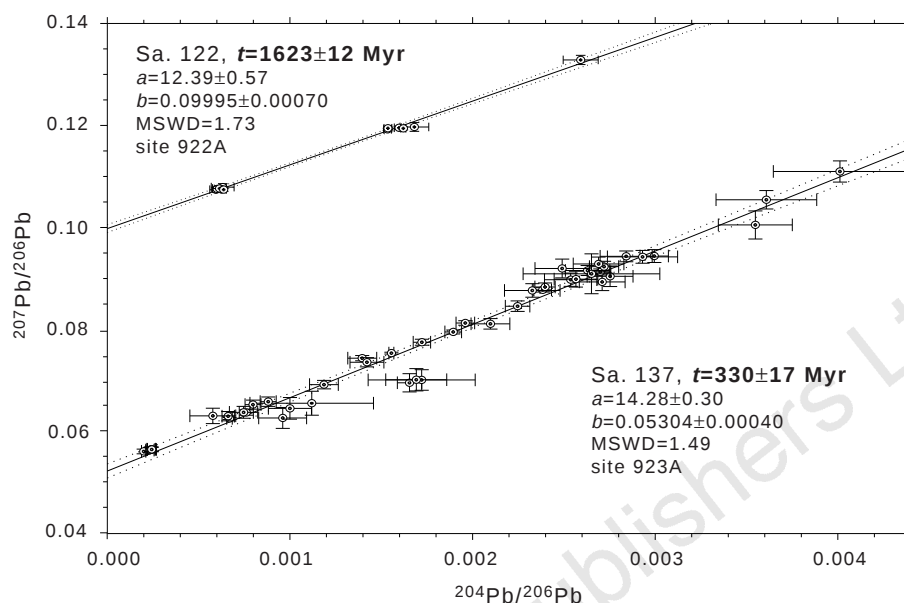


Figure 3 Plot of measured $^{207}\text{Pb}/^{206}\text{Pb}$ against measured $^{204}\text{Pb}/^{206}\text{Pb}$, used to obtain the $^{207}\text{Pb}/^{206}\text{Pb}$ age of single zircon groups. We used the mass-spectrometric evaporation method⁵. Every point represents a measuring block of 10 Pb isotope scans. 23 zircons of sample 137 (gabbro, site 923, hole A, ODP sample 10R-3/2A, 39.4 m below sea floor) and sample 134 (olivine gabbro, site 923, hole A, ODP sample 7R-1/2B, 27.0 m below sea floor) with 46 measuring blocks give the lower straight regression line (the zircons in sample 137 contain different, sometimes small amounts of lead): a is the slope of the regression line and MSWD indicates mean square of weighted deviates. The resulting radiogenic $^{207}\text{Pb}/^{206}\text{Pb}$

ratios b (at $^{204}\text{Pb}/^{206}\text{Pb} = 0$, calculated by York-fit³⁵ considering the individual errors of every point) corresponds to an age of 330 ± 17 Myr (95% confidence level). MSWD = 1.49. The lower line is essentially determined by 22 zircons of sample 137, but the 5 points of sample 134 of the same borehole situated beneath the $^{207}\text{Pb}/^{206}\text{Pb}$ axis plot extremely close to the line of sample 137 that indicates the same age for both samples. Two zircons of sample 122 (troctolite, site 922, hole A, ODP sample 2R-2/1C, 8.6 m below sea floor) contained abundant lead supplying 11 measuring blocks. These data give the upper regression line. The $^{207}\text{Pb}/^{206}\text{Pb}$ age is $1,623 \pm 12$ Myr.

1 Myr. Also, such a young non-equilibrium age of 0.15 Myr contradicts the age estimation from magnetic data suggesting a minimum age of 0.7–0.9 Myr (ref. 1).

For a further proof of the ages found, we performed conventional U–Pb dating (using the HF vapour digestion method) of 37 zircons (6 groups) from sample 137, and of 4 zircons (3 groups) additionally separated from sample 122. The results (Fig. 4; Table 1) fully support the evaporation ages that we obtained: sample 137 gives an upper intercept age of 301 Myr, which is identical to our 330-Myr evaporation age considering the errors. All zircons except one are strongly discordant. The least discordant zircons produced $^{206}\text{Pb}/^{238}\text{U}$, $^{207}\text{Pb}/^{235}\text{U}$ and $^{207}\text{Pb}/^{206}\text{Pb}$ ages of respectively 281 ± 2 , 284 ± 7 and 312 ± 67 Myr (2σ errors). Sample 122 b (second separation) has an upper intercept age of 1,722 Myr (least discordant ages of respectively $1,475 \pm 11$, $1,545 \pm 13$ and $1,658 \pm 12$ Myr), slightly higher than our evaporation age. The lower intercept age may coincide with the end of the Pan-African event; but in the possible case of continuous Pb loss this age has no meaning. First SHRIMP datings, each with two spots, resulted in an age of ~ 375 Myr for a zircon from sample 137 and an age of 1,650 Myr for a sample 122 zircon (first separation), confirming the earlier age determinations.

The discordance of our zircons suggests lead loss. Metzger and Krogstad²⁰ stated that lead diffusion in the pristine lattice is insignificant up to temperatures of at least 1,000 °C. Lee *et al.*²¹ showed that the closure temperature of the Th–U–Pb system in natural zircons is >900 °C. These workers determined lead diffusion coefficients at 900, 1,000 and 1,100 °C, obtaining mean values of respectively 5.1×10^{-21} , 6.8×10^{-19} and $9.1 \times 10^{-17} \text{ cm}^2 \text{ s}^{-1}$. Consequently, 100- μm zircons can retain their ‘memory’ at 900 °C for some 150 Myr; that is the time since the opening of the Atlantic Ocean. (Kresten *et al.*²² examined some zircons from

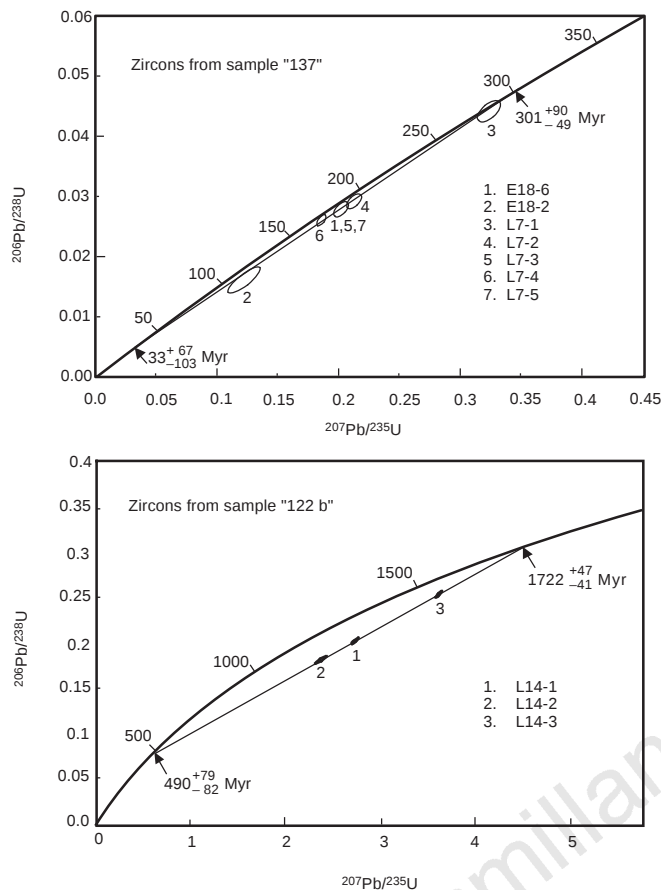
kimberlite and found unit-cell volumes 5% smaller than in normal zircons; this would diminish the Pb diffusion and raise their closure temperature).

The zircons that we studied were discovered in magmatic rocks which were drilled, not dredged. A direct continental influence must therefore be excluded, as the sample location is far away from the continental margins. Where do the zircons come from?

The occurrence of zircons in mantle-derived rocks is known from kimberlites^{22–27}. Kinny *et al.*²⁴ examined Jwang DK2 kimberlites from Botswana and found two populations of zircons: one was 235 Myr old, reflecting the eruption age; the other was 2,100–2,800 Myr old, showing the persistence of zircon xenocrysts originating in mantle-diamond fields. We also note that zircons in lower-crust and upper-mantle xenoliths carried in kimberlites are found to be substantially older than the kimberlite-eruption age²⁷.

Great ages have also been reported^{28–31} from other locations on the MAR—partly from continental type rocks. A granite-gneiss has been dredged at the MAR at 26° N, a region believed to be clearly outside the influence of glacial periods²⁹. From the recovered zircons U–Pb ages were found to be partly strongly discordant, the upper concordia–discordia intersection giving ages of 1,650 and 1,850 Myr; a whole-rock garnet Sm–Nd isochrone corresponded to an age of $1,687 \pm 15$ Myr (biotite K–Ar-age, $1,631 \pm 106$ Myr; ref. 31).

Bonatti *et al.*²⁸ showed that old sedimentary rocks also occur in the vicinity of the MAR. The 400-m-thick Romanche Sedimentary Sequence is situated ~ 100 km east of the eastern MAR–Romanche transform intersection. The lower unit contains dolomitized limestones bearing a microfossil population that marks an age of 140 Myr. Bonatti and co-workers²⁸ believe that this sedimentary sequence was deposited in a palaeo-Romanche continent–continent transform during the opening of the Atlantic Ocean. That the



present location of this sequence, in spite of sea-floor spreading, is near the MAR rather than the border of Africa is due to oscillatory spreading (repeated ridge jumping and transform migration)^{8,9}.

We now consider two possibilities for the origin and the path of our old zircons.

First, as seen above, zircons can reside in shallow regions of the upper mantle²⁴. But the established diffusion coefficients²¹ make it very improbable that these zircons could survive for >150 Myr in environments with temperatures in excess of 1,000 °C without totally losing their radiogenic lead; but times of the order of Myr are possible.

Rifting as a first step to the opening of the northern low-latitude Atlantic may result in upshearing of slices or slivers of the crust; it may also partially result in delamination of continental lithosphere fragments. Drifting develops the MAR and initiates upwelling of mantle material beneath it. Both movements drive small, roll-like, flat, slower-circulating cells^{6,7}. The cold detached slices and slivers sink: they could be taken up in such cells and—warmed and partly resorbed—transported to the vicinity beneath the MAR axis. There the remains, including the zircons, could be incorporated in the forming gabbros which—as a consequence of the ridge-transform dynamic—were tectonically uplifted^{32,34}. Alternatively, a delaminated slice may represent the harzburgite-gabbro unit totally or in part³⁰.

The second possible explanation may be a version of Bonatti's suggestion of repeated ridge jumping and transform migration^{8,9,34}. Sea-floor rock moves in opposite directions at the north and south

side of a transform fracture between the two intersections of the offset MAR with the transform fracture. Consequently, beneath the transform fracture there must be a non-drifting segment where older material could remain for some time. Moreover, the structure of transform fractures can change³² and migrate⁸: upsheared continental rock blocks⁹ could be transported by drift westwards immediately after the ocean opening, reside for some time in the slowly southward-moving overriding transform-structure, and drift eastwards to the area of gabbro formation. Due to the depth they could be more or less resorbed, and the remains—with the zircons—incorporated in the forming gabbros. □

Received 21 October 1997; accepted 19 March 1998.

1. Cannat, M. & ODP shipboard Scientific Party Leg 153 Probing the foundation of the Mid-Atlantic Ridge. *Eos* **76**, 129–133 (1995).
2. Werner, C.-D. Data report: geochemistry of rocks and minerals of the gabbro complex from the MARK area. *Proc. ODP Sci. Res.* **153**, 491–504 (1997).
3. Werner, C.-D. & Pilot, J. Data report: geochemistry and mineral chemistry of ultramafic rocks from the Kane area (MARK). *Proc. ODP Sci. Res.* **153**, 457–470 (1997).
4. Pilot, J., Werner, C.-D. & Tichomirow, W. Further oxygen isotope investigations from an ultramafic profile, hole 920D: alterations by seawater. *Terra Nostra* **4**, 66 (1996).
5. Kober, B. Single-zircon evaporation combined with Pb²⁰⁶ emitter bedding for ²⁰⁷Pb/²⁰⁶Pb-age determinations using thermal ion mass spectrometry, and implications to zirconology. *Contrib. Mineral. Petrol.* **96**, 63–71 (1987).
6. Nicolas, A. Die Ozeanischen Rücken 140, Fig. 7.7 (Springer, Berlin, 1995).
7. Rabinowicz, M., Nicolas, A. & Vigneresse, J. L. A rolling mill effect in asthenosphere beneath oceanic spreading centers. *Earth Planet. Sci. Lett.* **67**, 97–108 (1984).
8. Bonatti, E. & Crane, K. Oscillatory spreading explanation of anomalously old uplifted crust near oceanic transforms. *Nature* **300**, 343–345 (1982).
9. Kepezhinskas, P. & Dmitiev, D. Continental lithospheric blocks in central Atlantic Ocean. *Ophiolite* **17**, 19–35 (1992).
10. Cannat, M., Bideau, D. & Bougault, H. Serpentinized peridotites and gabbros in the Mid-Atlantic Ridge axial valley at 15° 37' N and 16° 52' N. *Earth Planet. Sci. Lett.* **109**, 87–106 (1992).
11. Karson, J. A. & Winters, A. T. in *Ophiolites and their Modern Oceanic Analogues* (eds Parsons, L. M., Murton, B. J. & Browning, P.) 107–116 (Spec. Publ. 60, Geol. Soc., Boulder, 1992).
12. Pupin, J. P. Zircon and granite petrology. *Contrib. Mineral. Petrol.* **73**, 207–220 (1980).
13. Gaggero, L. & Gazzotti, M. Primary and secondary oxides, sulfides and accessory minerals in Mid-Atlantic gabbros: mineralogy and petrology. *Ophiolite* **21**, 105–116 (1996).
14. Kelley, D. S. & Malpas, J. Melt-fluid evolution in gabbroic rocks from Hess Deep. *Proc. ODP Sci. Res.* **147**, 213–226 (1996).
15. Cannat, M., Ceuleneer, G., Fletscher, J. Localization of ductile strain and the magmatic evolution of gabbroic rocks drilled at the Mid-Atlantic Ridge (23° N). *Proc. ODP Sci. Res.* **153**, 84 (1997).
16. Tera, F. & Wasserburg, G. J. U-Th-Pb systematics in lunar highland samples from the Luna 20 and Apollo 16 mission. *Earth Planet. Sci. Lett.* **17**, 36–51 (1972).
17. Wendt, I. Radiometric methods in Geochronology. *Clausthaler Tektonische Hefte* **23**, 92–96 (1986). (In German.)
18. Ludwig, K. R. Effect of initial radioactive-daughter disequilibrium on U-Pb isotope apparent ages of young minerals. *J. Res. US Geol. Surv.* **5**, 663–667 (1977).
19. Wendt, I. & Carl, C. U/Pb dating of discordant 0.1 Ma old secondary U minerals. *Earth Planet. Sci. Lett.* **73**, 278–284 (1985).
20. Metzger, K. & Krogstad, E. J. Interpretation of discordant U-Pb zircon ages: an evaluation. *J. Metamorphic Geol.* **15**, 127–140 (1997).
21. Lee, J. K. W., Williams, I. S. & Ellis, D. J. Pb, U, and Th diffusion in natural zircon. *Nature* **390**, 159–162 (1997).
22. Kresten, P., Fels, P. & Berggren, G. Kimberlitic zircons—a possible aid in prospecting for kimberlites. *Mineral. Deposita (Berl.)* **10**, 47–56 (1975).
23. Ahrens, L. H., Cherry, R. D. & Erlank, A. J. Observations on the relationship in zircons from granitic rocks and from kimberlites. *Geochim. Cosmochim. Acta* **31**, 2379–2387 (1967).
24. Kinny, P. D., Compston, J., Bristow, J. W. & Williams, I. S. in *Kimberlites and Related Rocks* Vol. 2 (eds Ross, J. et al.) 833–842 (Geol. Soc. Austr. Spec. Publ. 14, Blackwell, Melbourne, 1989).
25. Jochum, K. P., Hofmann, A. W., Ito, E., Seufert, H. M. & White, W. M. K. U and Th in mid-ocean ridge basalt glasses and heat production, K/U and K/Rb in the mantle. *Nature* **306**, 431–436 (1983).
26. Hofmann, A. W. Chemical differentiation of the Earth: the relationship between mantle, continental crust, and oceanic crust. *Earth Planet. Sci. Lett.* **90**, 297–314 (1988).
27. Chen, Y. D., O'Reilly, S. Y., Kinny, P. D. & Griffin, W. L. Dating lower crust and upper mantle events: an ion microprobe study from kimberlitic pipes, South Australia. *Lithos* **32**, 77–94 (1994).
28. Bonatti, E. et al. Lower Cretaceous deposits trapped near the equatorial Mid-Atlantic Ridge. *Nature* **380**, 518–520 (1996).
29. Beljatsky, B. V. et al. Precambrian granitic gneiss from MAR (26° N): results of U-Pb and Sm-Nd isotopic investigations. *Geomija* **8**, 876–880 (1997). (In Russian.)
30. Truchalev, A. I. et al. Old K-Ar-age of a metagabbro and a granite-gneiss dredged at the axial part of the Mid-Atlantic Ridge, 26° N. *Dokl. Akad. Nauk* **311**, 1447–1452 (1990). (In Russian.)
31. White, R. & McKenzie, D. Magmatism at rift zones: the generation of volcanic continental margins and flood basalts. *J. Geophys. Res.* **94**(6), 7685–7729 (1989).
32. Tucholke, B. E. & Schouten, H. Kane fracture zone. *Mar. Geophys. Res.* **10**, 1–39 (1988).
33. Ludwig, K. R. *ISOPLOT—a Plotting and Regression Program for Radiogenic-Isotope Data V2.71* (Open File Rep. 91-445, US Geol. Surv., Denver, 1992).
34. Cannat, M. et al. Ultramafic and gabbroic exposures at the Mid-Atlantic Ridge: geological mapping in the 15° N region. *Tectonophysics* **279**, 193–213 (1997).
35. York, D. Least squares fitting of a straight line with correlated errors. *Earth Planet. Sci. Lett.* **5**, 320–324 (1969).

Acknowledgements. We thank P. Herzig and S. Petersen for comments; A. W. Hofmann, B. Hanan, F. Tera and S. Uhlir for discussions; U. Kempe for help with the SEM images; W. Todt (MPI Mainz) for supporting the conventional U-Pb dating; and I. S. Williams (ANU, RSES, Canberra) for SHRIMP datings. This work was supported by the Deutsche Forschungsgemeinschaft.

Correspondence and requests for materials should be addressed to J.P. (e-mail: pilot@mineral.tu-freiburg.de).

A king-sized theropod coprolite

Karen Chin, Timothy T. Tokaryk*, Gregory M. Erickson†‡ & Lewis C. Calk

United States Geological Survey, 345 Middlefield Road, MS 975, Menlo Park, California 94025, USA

* Eastend Fossil Research Station, Royal Saskatchewan Museum, Box 460, Eastend, Saskatchewan S0N 0T0, Canada

† Department of Integrative Biology & Museums of Vertebrate Zoology and Paleontology, University of California, Berkeley, California 94720, USA

Fossil faeces (coprolites) provide unique trophic perspectives on ancient ecosystems. Yet, although thousands of coprolites have been discovered, specimens that can be unequivocally attributed to carnivorous dinosaurs are almost unknown. A few fossil faeces have been ascribed to herbivorous dinosaurs^{1–3}, but it is more difficult to identify coprolites produced by theropods because other carnivorous taxa coexisted with dinosaurs and most faeces are taxonomically ambiguous. Thus sizeable (up to 20 cm long and 10 cm wide) phosphatic coprolites from Belgium⁴ and India^{5,6} that have been attributed to dinosaurs might have been produced by contemporaneous crocodylians⁷ or fish. But there is no ambiguity about the theropod origin of the Cretaceous coprolite we report here. This specimen is more than twice as large as any previously reported carnivore coprolite, and its great size and temporal and geographic context indicate that it was produced by a tyrannosaur, most likely *Tyrannosaurus rex*. The specimen contains a high proportion (30–50%) of bone fragments, and is rare tangible evidence of theropod diet and digestive processes.

The specimen (SMNH P2609.1) was discovered as an elongate mass weathering out of the fluvial Maastrichtian Frenchman Formation in Southwestern Saskatchewan, roughly 35 km southeast of the town of Eastend. The fractured mass was distinguished by its indurated nature and numerous inclusions of comminuted bone. The main portion of the mass was found *in situ* in a bentonitic mudstone, though numerous fragments had eroded downslope. No fossil bones were found in association with the coprolite, but fossils of a number of large vertebrates have been recovered from the Frenchman Formation⁸.

The main body of the specimen is roughly 44 cm long, 13 cm high and 16 cm wide (Fig. 1). The density of the material (approximately 2.94 g ml⁻¹) and the weight of all portions (over 7.1 kg) indicate that the present volume of the mass is ~2.4 litres, though it is likely that the original faecal mass was larger before it was subjected to compaction, attrition, and/or desiccation. Broken surfaces of the specimen expose numerous dark brown macroscopic bone fragments ranging from 2 to 34 mm in length. These pieces are suspended in a microcrystalline ground mass and are generally aligned in a consistent direction. The ground mass also contains sand-sized bone clasts (Fig. 2). Most of the included bone appears to be similar in type, with highly vascularized cortical bone tissue up to 14-mm thick in a fibrolamellar pattern. All of the observed bone is primary, and no lines of arrested growth were detected.

Bulk chemical analyses using X-ray fluorescence (Table 1) reveal marked differences between the specimen and the Frenchman Formation mudstone. The bone-bearing specimen contains high concentrations of phosphorus and calcium, and lower concentrations of aluminium and silicon, relative to the host sediment. Microprobe analyses of specific areas of the specimen indicate that the bone fragments and coprolitic ground mass have similar

compositions, though the ground mass contains more silicon and aluminium (Table 2). X-ray powder-diffraction analyses indicate that carbonate fluorapatite is the predominant phosphate mineral in both the bone and the ground mass.

Several factors confirm that this specimen is a coprolite. The most diagnostic feature is a phosphatic composition, which is characteristic of carnivore coprolites⁹. As phosphorus normally constitutes only about 0.1% of the Earth's crustal rocks¹⁰, concentrated phosphate deposits usually indicate biotic accumulations, and the overall configuration of the mass is consistent with the irregular faecal deposits produced by very large animals. The matrix-supported distribution of bone fragments argues against the possibility that the mass represents regurgitated material or fluvially aggregated bone debris.

The tremendous size of the specimen indicates that the faecal mass was produced by a large theropod. The largest theropod found in the Frenchman Formation is *Tyrannosaurus*, with an estimated body weight of 5,400 to 6,300 kg (ref. 11). Although other theropods, crocodylomorphs, and a chelonian (*Dromaeosaurus*, *Saurornitholestes*, *Troodon*, *Richardoestesia*, *Chirostenotes*, an ornithomimid, *Leidyosuchus*, *Champsosaurus*, and *Aspideretes*) have also been recovered from the Frenchman Formation⁸, these smaller carnivorous taxa probably weighed <100 kg (ref. 12), and it



Figure 1 Large, bone-bearing theropod coprolite with some of the broken pieces that had eroded downslope. This specimen was found in Chamberly Coulee in the Frenchman River Valley, roughly 11.5 m below the Cretaceous/Tertiary boundary. Scale bar, 10 cm.

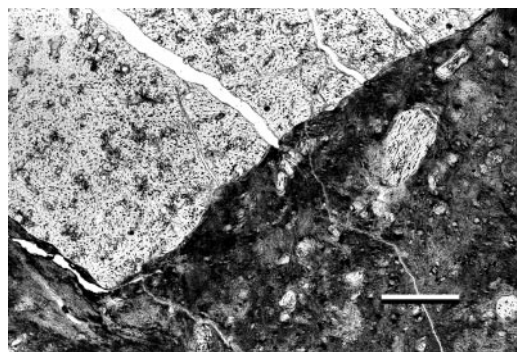


Figure 2 Photomicrograph of a thin section of the theropod coprolite, showing sand- to pebble-sized bone clasts within a microcrystalline phosphatic ground mass. The elemental composition of the ground mass is similar to that of the bone fragments, indicating that it is probably largely composed of reprecipitated bone apatite infiltrated by clay minerals from the host sediment (Table 1). The large bone fragment in the upper left portion of the image exhibits a fibrolamellar pattern, with osteocyte lacunae concentrically arranged around the vascular canals. Probe measurements of the interior of bone lacunae revealed that many of these channels are at least partially empty, whereas others exhibit variable element distributions, with generally lower concentrations of calcium and phosphorus, and higher silicon and aluminium levels (Table 2). Scale bar, 400 µm.

‡ Present address: Biomechanical Engineering Division, Mechanical Engineering Department & Rehabilitation R & D Center, VA Health Care System, Stanford University, Stanford, California 94305, USA.

is unlikely that they could have produced large quantities of faeces. The mass could have been produced by a different species of tyrannosaur, but no others have been recognized in the Frenchman Formation.

The stomach acids and proteolytic enzymes of large extant carnivores digest bone to varying degrees^{13–16}. Tangible evidence of this process is apparent in areas of the theropod coprolite where aligned and rounded bone pieces represent the degraded remains of large bone fragments (Fig. 3). The contents of carnivore coprolites might reflect animal physiology, because the extent of bone digestion can be indicative of gut-residence time¹⁷. Carnivore feeding activity is usually variable, however, and stomach acidity and gut-residence time can be altered by non-physiological factors such as frequency of meals¹⁵. Even so, the high percentage of incompletely digested bone in this specimen is interesting because it is inconsistent with the general prediction that large theropods digested most consumed bone¹⁸ in the manner of extant crocodilians¹⁴.

The chemistry of the coprolite reflects several factors. A large percentage of the phosphate of the ground mass was probably derived from dissolved bone apatite, but other dietary residues would have contributed additional phosphorus, as microorganisms and animal soft tissues contain significant concentrations of this element. Postdepositional phosphate precipitation may have been triggered by bacterial enzymes¹⁹ after burial of the faecal mass. Although the chemistry of this diagenetic phosphate is similar to that of the included bone, the increased amounts of silicon and aluminium and small differences in amounts of other elements

fragments give clues to the identity of the ingested animal. Dinosaurs are the only Late Cretaceous animals that regularly produced thick fibrolamellar cortical bone²⁰. The absence of secondary osteons indicates that the bone was ontogenetically juvenile, so the ingested animal appears to have been a subadult dinosaur. Although bone histology is not species-specific, the absence of arrested growth lines may indicate an ornithischian dinosaur. Lines of arrested growth have been observed in several theropods²¹ but have not been observed in the long bones of *Triceratops* and *Edmontosaurus* (G.M.E., unpublished observation), the most common dinosaurs found in the Frenchman Formation. Other ornithischians from the formation include *Torosaurus*, *Thescelosaurus* and an ankylosaur⁸.

The thickness of the cortical bone indicates that the fragments may be derived from appendicular bone or ceratopsian frill. If the fragments were derived from long-bone diaphyses, estimates of the weight²² of the consumed animal might range from ~200 kg (for a bipedal dinosaur) to 750 kg (for a quadrupedal dinosaur). The pronounced fragmentation and angularity of the consumed bone indicate that it was fractured before ingestion—apparently by biting during feeding. Although extant birds (avian dinosaurs) often use a horny gizzard and/or ingested grit for food maceration²³, such mechanisms could not have been solely responsible for the degree of comminution seen in the coprolite. Furthermore, significant trituration would have resulted in well rounded bone clasts, and there is no evidence for the use of gastroliths by non-avian theropods.

The coprolitic evidence for extensive bite-induced bone fragmentation is surprising in view of modern reptilian feeding habits. Extant reptiles have poor dental occlusion and generally swallow

Table 1 X-ray-fluorescence data of the weight percentage of oxides found in bulk powdered samples of the coprolite and host sediment

	Coprolite	
SiO ₂	7.93	70.2
Al ₂ O ₃	2.43	18.1
TiO ₂	0.128	0.794
FeO	0.95	6.02
MnO	0.151	0.018
CaO	44.6	0.97
MgO	0.16	1.42
K ₂ O	0.27	2.22
Na ₂ O ₂	0.28	1.08
P ₂ O ₅	26.5	0.70
Total	83.5	100.9

The clay-rich sediment reflects the low phosphorus concentration of most inorganic rocks, whereas the coprolite is largely composed of biotically concentrated phosphorus and calcium.

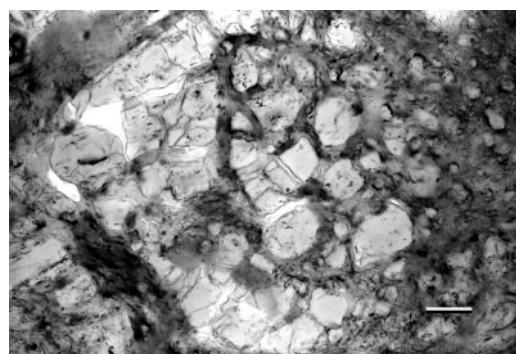


Figure 3 Photomicrograph of a thin section of the theropod coprolite, showing associated bone fragments that indicate digestive degradation. Digestive acids and enzymes probably infiltrated the interior of the bone through vascular canals. Scale bar, 100 μm.

Table 2 Microprobe data indicating weight percentage of oxides, fluorine and chlorine in different regions of the coprolite

	Bone fragments n = 60 points		Ground mass n = 60 points		Bone lacunae n = 27 points	
	Mean	s.d.	Mean	s.d.	Mean	s.d.
SiO ₂	0.038	0.102	4.03	1.42	29.5	13.2
Al ₂ O ₃	0.094	0.046	1.71	0.516	13.6	5.80
FeO	0.484	0.034	0.730	0.101	3.24	1.39
MnO	0.230	0.026	0.144	0.018	0.068	0.046
MgO	0.103	0.018	0.210	0.030	1.02	0.47
CaO	51.5	0.290	48.9	1.09	20.2	13.5
SrO	0.157	0.030	0.117	0.024	0.036	0.040
Na ₂ O	0.335	0.046	0.237	0.032	0.184	0.056
K ₂ O	0.025	0.010	0.163	0.044	0.312	0.110
SO ₃	0.142	0.055	0.285	0.029	0.157	0.080
P ₂ O ₅	35.7	0.466	32.6	0.859	13.4	9.37
F	2.95	0.074	2.62	0.117	0.943	0.750
Cl	NA	—	NA	—	0.108	0.090
Total	90.5	0.582	90.7	0.584	82.4	4.96

The compositions of the bone and ground mass are similar, though the ground mass appears to contain more contributions from the host sediment. Of 67 probe measurements of lacunae, 40 registered low element totals, indicating that the vascular canals were incompletely filled. These channels would have acted as conduits for digestive fluids and for postdepositional contaminants. Data listed above are from the 27 lacunae that registered element totals over 70%. NA, no analyses done.

large pieces of prey whole^{14,24}. Such observations of modern feeding behaviours have led to speculation that extinct theropods did little bone-crushing^{18,25} and wasted a significant proportion of the food available from carcasses²⁶. Tyrannosaur teeth appear to be stout enough to damage bone²⁷, however, and analyses of bite marks on *Triceratops* and *Edmontosaurus* bones indicate that *Tyrannosaurus* pulverized bones during feeding²⁸ and probably consumed bone fragments²⁹.

Although a single coprolite cannot be construed as representative of diet, this rare example of fossilized dietary residues helps to refine our understanding of theropod feeding behaviour by providing physical evidence that a tyrannosaur crushed, consumed, and incompletely digested large quantities of bone when feeding on a subadult dinosaur. It also presents a new search image for future discoveries of theropod faeces that will help us to elucidate the food habits of these giant meat-eaters. □

Methods

Bulk chemical analyses of the coprolite and host sediment (Table 1) were made on a Rigaku 3370 spectrometer by the staff of the GeoAnalytical Laboratories, Washington State University, using described procedures³⁰.

Mineralogical compositions of the bone and ground mass were examined with a Phillips V2.0 diffractometer at the University of California, Santa Barbara (scanning from 2–80° 2 θ). Elemental analyses of these components (Table 2) were made on a JEOL 8900 microprobe at the US Geological Survey, Menlo Park. A 15 kV, 20 nA beam defocused to produce a spot size of 15 μ m was used to analyse bone and ground mass; a 10 nA current was used for lacunae analyses. Natural minerals (Wilberforce apatite, Tiburon albite, strontianite, barite, San Carlos olivine, and sodalite) and synthetic materials (faylite, Mn₂O₃, An100, and GSC glass) were used as standards.

Received 22 January; accepted 23 March 1998.

- Hill, C. R. Coprolites of *Ptiliophyllum* cuticles from the Middle Jurassic of North Yorkshire. *Bull. Br. Mus. Nat. Hist.* **27**, 289–294 (1976).
- Chin, K. & Gill, B. D. Dinosaurs, dung beetles, and conifers: participants in a Cretaceous food web. *Palaios* **11**, 280–285 (1996).
- Chin, K. & Kirkland, J. I. Probable herbivore coprolites from the Upper Jurassic Mygatt-Moore Quarry, Western Colorado. *Mod. Geol.* **23**, 249–276 (1998).
- Bertrand, C. E. Les Coprolithes de Bernissart. I. partie: Les Coprolithes qui ont été attribuées aux Iguanodons. *Royal Mus. Hist. Nat. Belgique Mem.* **1**, 1–154 (1903).
- Matley, C. A. The coprolites of Pijidura, Central Provinces. *Geol. Surv. Recs.* **74**, 535–547 (1939).
- Jain, S. L. in *Dinosaur Tracks and Traces* (eds Gillette, D. D. & Lockley, M. G.) 99–108 (Cambridge Univ. Press, Cambridge, UK, 1989).
- Abel, O. Diskussion zu den Vorträgen R. Kräusel und F. Versluys. *Palaeontologische Zeitschrift* **4**, 87 (1922).
- Tokaryk, T. T. in *Canadian Paleontology Conference Field Trip Guidebook No. 6* (ed. McKenzie-McAnally, L.) 34–44 (Geol. Assoc. Canada, St John's, Newfoundland, 1997).
- Hunt, A. P., Chin, K. & Lockley, M. G. in *The Palaeobiology of Trace Fossils* (ed. Donovan, S. K.) 221–240 (Wiley, Chichester, 1994).
- Stienstra, P. *Sedimentary Petrology, Origin and Mining History of the Phosphate Rocks of Klein Curacao, Curacao and Aruba, Netherlands West Indies* No. 130, 207 (Pub. Found. Sci. Res. Caribbean Region, 1991).
- Farlow, J. O., Smith, M. B. & Robinson, J. M. Body mass, bone “strength indicator”, and cursorial potential of *Tyrannosaurus rex*. *J. Vert. Paleontol.* **15**, 713–725 (1995).
- Peczis, J. Implications of body-mass estimates for dinosaurs. *J. Vert. Paleontol.* **14**, 520–533 (1994).
- Duke, G. E., Jegers, A. A., Loff, G. & Evanson, O. A. Gastric digestion in some raptors. *Comp. Biochem. Physiol.* **A 50**, 649–659 (1975).
- Fisher, D. C. Crocodilian scatology, microvertebrate concentrations, and enamel-less teeth. *Paleobiology* **7**, 262–275 (1981).
- Andrews, P. *Owls, Caves and Fossils* 231 (Univ. Chicago Press, Chicago, 1990).
- Denys, C., Fernandez-Jalvo, Y. & Dauphin, Y. Experimental taphonomy: preliminary results of the digestion of micromammal bones in the laboratory. *Comptes Rendus Acad. Sci.* **321**, 803–809 (1995).
- Mellet, J. S. Dinosaurs, mammals and Mesozoic taphonomy. *Acta Palaeontol. Pol.* **28**, 209–213 (1983).
- Hunt, A. P. Phanerozoic trends in nonmarine taphonomy: implications for Mesozoic vertebrate taphonomy and paleoecology. *Geol. Soc. Am. Abstr.* **19**, 171 (1987).
- Lucas, J. & Prévôt, L. E. in *Taphonomy: Releasing the Data Locked in the Fossil Record* (eds Allison, P. A. & Briggs, D. E. G.) 389–409 (Plenum, New York, 1991).
- de Ricqlès, A. J. in *Morphology and Biology of Reptiles* (eds Bellairs, A. d'A. & Cox, C. B.) 123–150 (Academic, London, 1976).
- Reid, R. E. H. Zonal “grown rings” in dinosaurs. *Mod. Geol.* **15**, 19–48 (1990).
- Anderson, J. F., Hall-Martin, A. & Russell, D. A. Long-bone circumference and weight in mammals, birds and dinosaurs. *J. Zool. A* **207**, 53–61 (1985).
- Stevens, C. E. & Hume, I. D. *Comparative Physiology of the Vertebrate Digestive System* 2nd edn, 400 (Cambridge Univ. Press, Cambridge, UK, 1995).
- Auffenberg, W. *The Behavioral Ecology of the Komodo Monitor* 406 (Univ. Presses of Florida, Gainesville, 1981).
- Fiorillo, A. R. Prey bone utilization by predatory dinosaurs. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **88**, 157–166 (1991).
- Farlow, J. O. A consideration of the trophic dynamics of a Late Cretaceous large-dinosaur community (Oldman Formation). *Ecology* **57**, 841–857 (1976).

- Farlow, J. O., Brinkman, D. L., Abler, W. L. & Currie, P. J. Size shape and serration density of theropod dinosaur lateral teeth. *Mod. Geol.* **16**, 161–198 (1991).
- Erickson, G. M. et al. Bite-force estimation for *Tyrannosaurus rex* from tooth-marked bones. *Nature* **382**, 706–708 (1996).
- Erickson, G. M. & Olson, K. H. Bite marks attributable to *Tyrannosaurus rex*: preliminary description and implications. *J. Vert. Paleontol.* **16**, 175–178 (1996).
- Hooper, P. R., Johnson, D. M. & Conrey, R. M. Major and trace element analyses of rocks and minerals by automated X-ray spectrometry (Washington State Univ., Geol. Dept. Open File Report, 1993).

Acknowledgements. We thank W. Sloboda for the discovery of the coprolite; the Allemand family for permitting access to the land and its fossils; H. N. Bryant, M. J. Jurashius, C. E. Meyer, M. Moreno, R. L. Oscarson, J. F. Parham, D. Pierce, J. Rifkin and B. H. Tiffany for comments and technical assistance; the Royal Saskatchewan Museum, Stanford University, the US Geological Survey, and the University of California at Santa Barbara for technical and administrative assistance; and the late W. V. Sliter for support and encouragement.

Correspondence and requests for materials should be addressed to K.C. (e-mail: kchin@pangea.stanford.edu).

Genetics underlying inbreeding depression in *Mimulus* with contrasting mating systems

Michele R. Dudash & David E. Carr*

Department of Biology, University of Maryland, College Park, Maryland 20742, USA

*Blandy Experimental Farm, University of Virginia, Route 2, Box 210, Boyce, Virginia 22602, USA

The importance of inbreeding depression in theoretical considerations of mating-system evolution^{1–5} and its potential impact on the persistence of small populations⁶ has renewed interest in the genetic basis of this phenomenon. Inbreeding increases homozygosity. This can produce inbreeding depression for two different reasons: first, deleterious recessive or partially recessive alleles that are masked at heterozygous loci by dominant alleles become fully expressed in homozygotes; and second, alleles may interact in an overdominant manner, such that the fitness of either type of homozygote is lower than that of heterozygotes. These two mechanisms produce different long-term effects in populations experiencing increased levels of inbreeding. Inbreeding depression resulting from deleterious alleles can be removed by selection, but inbreeding depression produced by overdominance cannot be removed without lowering the mean fitness of the population^{1–5}. Using a North Carolina 3 breeding programme⁷, the most powerful quantitative genetics technique available^{8–10}, we show here that deleterious recessive alleles are mainly responsible for inbreeding depression in two closely related annual plants, the primarily selfing *Mimulus micranthus* and the mixed-mating *M. guttatus*. Estimates indicate that deleterious alleles in *M. micranthus* are more nearly additive than they are in *M. guttatus*.

The genetic basis of inbreeding depression (or its converse, heterosis) has been examined primarily in crop plants. There is evidence for both dominance-based^{11–13} and overdominance-based^{13–16} inbreeding depression. However, the relative importance of dominance-based versus overdominance-based inbreeding depression in natural plant populations is largely unknown. Studies of *Eichhornia paniculata*¹⁷ (Pontederiaceae) and two *Amsinckia* species¹⁸ (Boraginaceae) have found indirect evidence for dominance-based inbreeding depression.

The genus *Mimulus* (Scrophulariaceae) has been the focus of many studies aimed at understanding the processes responsible for the evolution of plant mating systems^{19–24}. *Mimulus guttatus*, the common monkey flower, has large, bee-pollinated flowers, and measured outcrossing rates for three populations, including one used in this study, ranged from 0.68–0.80 (ref. 25). *Mimulus micranthus* is a

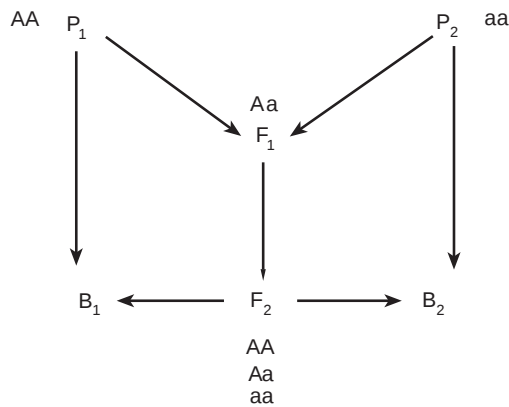


Figure 1 The North Carolina 3 breeding design. Starting with field-collected seed, inbred lines of *M. guttatus* were produced through five sequential generations of selfing in a glasshouse. Maternal lineages from selfing *M. micranthus* were allowed to self for three generations. Inbred lines (P_1 and P_2) were crossed to produce hybrid F_1 progeny. F_1 individuals were selfed to produce the F_2 generation, representing all possible assortments of linkage groups from the parental inbred lines as well as many recombinant genotypes. Backcross progeny (B_1 and B_2) were produced by crossing randomly selected F_2 pollen parents to each of the original inbred lines (P_1 and P_2).

small-flowered, primarily selfing derivative of *M. guttatus* with an estimated outcrossing rate of 0.16 (ref. 20). *M. guttatus* is found widely throughout western North America whereas *M. micranthus* is restricted to the Coastal Ranges of northern California. The magnitude of inbreeding depression in mixed-mating *M. guttatus* is much greater than in selfing *M. micranthus*²⁴ with respect to above-ground biomass, pollen production, and ovule production. Here we present estimates for the average dominance levels responsible for the inbreeding depression in these same traits as well as for total flower production and pollen viability.

Most fitness traits are probably affected by many loci. Some of these loci could be segregating alleles that act in an overdominant manner, whereas others could be segregating alleles that act in a dominant-recessive manner. We used the North Carolina 3 breeding design (Fig. 1) to estimate average dominance, a mean of the various allelic interactions across all loci affecting particular fitness traits⁷. Average dominance ($\bar{a}$) represents the ratio of non-additive genetic variance to additive genetic variance (Table 1). A ratio of more than 1 indicates inbreeding depression due to overdominance, whereas a ratio of less than 1 indicates inbreeding depression due to partially recessive deleterious alleles. In both *Mimulus* species, inbreeding depression in most life-history traits results from partially recessive deleterious alleles (Table 2). All estimates of $\bar{a}$ for total flower production, above-ground biomass, and ovule production were significantly less than 1.0 (ref. 26). For pollen viability in both populations of *M. guttatus*, $\bar{a}$ values did not differ significantly from 1.0 and are therefore indistinguishable from complete dominance of wild-type alleles. Despite the similarity in average dominance values for most traits between populations of *M. guttatus*, estimates of average dominance for pollen production differed in each population (Table 2). Population S showed partial dominance (approaching additivity; $\bar{a}$ significantly <1.0), but population T exhibited evidence of overdominance ($\bar{a}$ significantly >1.0). If deleterious alleles are linked in repulsion, estimates from the NC3 design are biased towards overdominance⁷. Our data therefore indicate that allelic interactions are not consistent and/or that the degree of linkage disequilibrium may vary among populations. Direct estimates of the genetics underlying inbreeding depression in other natural populations are needed.

Estimates of average dominance, $\bar{a}$, for total flower production

Table 1 Representative ANOVA of backcross progeny from a North Carolina 3 breeding design

Source of variation	d.f.	MS	Exp (MS)
Set	$s - 1$	—	—
Replication (set)	$s(r - 1)$	—	—
Inbred lines (set)	s	—	—
F_2 parents (set)	$s(n - 1)$	M_{31}	$Ve + r\alpha$
$F_2 \times$ line (set)	$s(n - 1)$	M_{32}	$Ve + r\delta$
Error	$s(2n - 1)(r - 1)$	M_{33}	Ve

Expected mean squares (Exp (MS)) are as presented in ref. 10, where Ve includes part of the genetic variance as well as the environmental variance. α , genetic variation among F_2 (male) parents (additive component); δ , interaction between F_2 genotypes and inbred parents (non-additive or dominance component); d.f., degrees of freedom; s , set of progeny, equivalent to the number of paired inbred lines mated to form the F_1 sets; r , number of progeny measured per cross, represents level of replication of each resultant backcross progeny set.

Table 2 Estimates of average dominance ($\bar{a}$) responsible for inbreeding depression in traits from two populations (S and T) of mixed-mating *M. guttatus* and primarily selfing *M. micranthus*

Trait	Average dominance		
	<i>M. guttatus</i>		<i>M. micranthus</i>
	S	T	
Total flower production	0.778*	0.741*	0.436***
Above-ground biomass	0.747*	0.692**	0.577***
Ovule production	0.613***	0.601***	—
Pollen viability	0.972	0.931	—
Pollen production	0.174***	1.321*	—

Significance levels follow the convention of * $P < 0.05$, ** $P < 0.01$, and *** $P < 0.001$ (ref. 26). No inbreeding depression was detected for ovule and pollen traits in *M. micranthus*²⁴.

and above-ground biomass in the inbred derivative, *M. micranthus*, are consistently lower (more nearly additive) than in its outbred ancestor, *M. guttatus*. Although the 95% confidence intervals do not allow us to distinguish statistically between the *M. micranthus* and the *M. guttatus* estimates¹⁰, models for purging genetic load¹⁻³ predict such differences. Selection against highly recessive alleles becomes much more efficient as selfing rates increase. Increased selfing rates will have little effect, however, on those alleles that are more nearly additive (dominance coefficient $h = 0.5$; $\bar{a} = 0$)².

Comparisons of species or populations that differ in their mating system indicate that inbreeding depression may be due to deleterious recessive alleles (presumably introduced by mutation)^{22,27}. This trend is also observed for *M. guttatus* and *M. micranthus*²⁴. We have now provided direct support for this long-standing inference with data from natural populations. As $\bar{a}$ is an estimate of the average dominance effects across all loci affecting a trait, we cannot preclude the action of overdominance at any particular locus. However, even after the inferred removal of deleterious alleles in highly selfing *M. micranthus*²⁴, we see no evidence for overdominance. Additional support for dominance-based inbreeding depression comes from a five-generation serial inbreeding and outcrossing program in *M. guttatus* that generated some maternal family lines in which the genetic load was successfully purged²⁸ for the same traits that we studied. A study of allozyme segregation patterns also showed little evidence for overdominance-based inbreeding depression²¹. Our intermediate levels of dominance for most life-history traits in *Mimulus* also concur with studies of two primarily selfing *Amsinckia* species¹⁸. One primary concern of conservation genetics is the loss of population fitness through inbreeding that is associated with reduced population size and fragmentation. Breeding programs aimed at rapidly purging a population of its inbreeding depression through inbreeding and selection will be successful only if deleterious recessive alleles are responsible⁶. Our results from a natural system indicate that such breeding programs may be worthwhile for some wild populations. □

Methods

We followed the outline of the North Carolina 3 breeding design described in ref. 7. To maximize the power of the design, we mated ten pairs of *M. guttatus* inbred lines in a negative assortative fashion with regard to overall vigour in

each of two populations (named S and T). We restricted the mating of ten pairs of *M. micranthus* inbred lines to those with fixed alternative isozyme markers, to allow confirmation of successful crosspollination in this autogamous selfer. The degree of dominance was quantified by partitioning variance measured in the backcross (B) generation. The B₁ and B₂ families (constructed crosses between $n = 20$ F₂ sires mated to each original inbred line, P₁ and P₂) make up one set of progenies (s), and we replicated each set $r = 5$ times. For each population of *M. guttatus* and for the single population of *M. micranthus*, ~2,000 backcross progeny grew in a randomized block design with B₁ and B₂ progeny randomized within each set and sets randomized within blocks.

The ratio in equation (1) estimates the square of the average dominance of wild-type alleles:

$$\frac{M_{32} - M_{33}}{M_{31} - M_{33}} = \bar{a}^2 \quad (1)$$

where M₃₁, M₃₂, and M₃₃ refer to the expected mean squares from the analysis of variance (ANOVA) in Table 1. This is the ratio of the interaction effect between F₂ pollen parents and the parental inbred lines to the additive genetic effects among F₂ pollen parents. If there is little or no dominance, the mean performance of progeny sired by a given F₂ pollen donor should be similar in both the B₁ and B₂ families. The pollen donor $\times$ inbred line interaction effect (the numerator in equation (1)) will be very small in this case, and the ratio will approach 0. With complete or partial dominance of wild-type alleles (that is, complete or partial recessivity of deleterious alleles), the mean performance of progeny sired by a given F₂ pollen parent will differ between the B₁ and B₂ families because of differences in heterozygosity between progeny of B₁ and B₂ families (Fig. 1). In this case, the magnitude of the interaction effect approaches the additive effect (as dominance becomes more complete), and the ratio approaches 1. With overdominance, the performance of the backcross progeny sired by a given F₂ pollen donor will be determined almost entirely by the interaction between the pollen donor and the inbred line to which it is crossed. The cross resulting in the most heterozygous progeny will outperform the other. In this case, almost all variation among backcross families is nonadditive, and the ratio is greater than 1.

The average dominance estimate, $\bar{a}$, is related to h , the dominance coefficient used in many population genetic models^{1,2}, in the following way: $h = 0.5 - \bar{a}/2$, with estimates in the range $0 > \bar{a} > 1.0$. When complete directional dominance is responsible for the genetic load, $h = 0$ and equivalently $\bar{a} = 1$. In the case of pure additivity, $h = 0.5$ and $\bar{a} = 0$. In the case of overdominance, $\bar{a} > 1$ and there is no equivalent value of h .

There are several assumptions⁷ for the genetic interpretation of the variance components listed in Table 1. Here the assumptions regarding linkage and epistasis are the most critical. Genetic interpretation of the ANOVA assumes that backcross progeny are in linkage equilibrium for all loci affecting the traits and that no nonallelic interactions occur. If deleterious recessive alleles are linked in repulsion, estimates of $\bar{a}$ are inflated because of pseudo-overdominance. Epistatic effects also upwardly bias estimates of $\bar{a}$, because the numerator in equation (1) will include nonallelic as well as allelic interactions. The bias caused by ignoring epistasis is minor compared with the potential bias caused by ignoring linkage disequilibrium. The few data available on the contribution of epistatic interactions to heterosis indicate that epistatic effects are minor compared with dominance effects^{19,28–30}.

Received 23 December 1997; accepted 15 April 1998.

- Charlesworth, D. & Charlesworth, B. Inbreeding depression and its evolutionary consequences. *Annu. Rev. Ecol. Syst.* **18**, 237–268 (1987).
- Charlesworth, D., Morgan, M. T. & Charlesworth, B. Inbreeding depression, genetic load, and the evolution of outcrossing rates in a multilocus system with no linkage. *Evolution* **44**, 1469–1489 (1990).
- Lande, R. & Schemske, D. W. The evolution of self-fertilization and inbreeding depression in plants. I. Genetic models. *Evolution* **39**, 24–40 (1985).
- Holsinger, K. E. Inbreeding depression and the evolution of plant mating systems. *Trends Ecol. Evol.* **6**, 307–308 (1991).
- Uyenoyama, M. K., Holsinger, K. E. & Waller, D. M. in *Oxford Surveys in Evolutionary Biology* Vol. 9 (eds Futuyma, D. & Antonovics, J.) 327–381 (Oxford Univ. Press, New York, 1993).
- Fenster, C. B. & Dudash, M. R. in *Restoration of Endangered Species* (eds Bowles, M. L. & Whelan, C. J.) 34–62 (Cambridge Univ. Press, Cambridge, UK, 1994).
- Comstock, R. E. & Robinson, H. F. in *Heterosis* (ed. Gowen, J. W.) 494–516 (Iowa State College Press, Ames, 1952).
- Jinks, J. L. in *Heterosis: Reappraisal of Theory and Practice* (ed. Frankel, R.) 1–46 (Springer, Berlin, 1983).
- Kearsey, M. J. The efficiency of the North Carolina Experiment III and the selfing backcrossing series for estimating additive and dominance variation. *Heredity* **45**, 73–82 (1980).
- Bulmer, M. G. *The Mathematical Theory of Quantitative Genetics* (Clarendon, Oxford, 1985).
- El-Nahrawy, M. A. & Bingham, E. T. Performance of S1 alfalfa lines from original and improved populations. *Crop Sci.* **29**, 920–923 (1989).

- Moll, R. H., Lindsey, M. F. & Robinson, H. F. Estimates of genetic variances and level of dominance in maize. *Genetics* **49**, 411–423 (1964).
- Hallauer, A. R. & Miranda, J. B. *Quantitative Genetics in Maize Breeding* (Iowa State Univ. Press, Ames, 1985).
- Apirion, D. & Zohary, D. Chlorophyll lethal in natural populations of the orchard grass (*Dactylis glomerata* L.). A case of balanced polymorphism in plants. *Genetics* **46**, 393–399 (1961).
- Williams, W. & Brown, A. G. Genetic response to selection in cultivated plants: gene frequencies in *Prunus avium*. *Heredity* **10**, 237–245 (1956).
- Gustafsson, A. The cooperation of genotypes in barley. *Heredity* **39**, 1–18 (1950).
- Barrett, S. C. H. & Charlesworth, D. Effects of a change in the level of inbreeding on the genetic load. *Nature* **352**, 522–524 (1991).
- Johnston, M. O. & Schoen, D. J. Mutation rates and dominance levels of genes affecting total fitness in two angiosperm species. *Science* **267**, 226–229 (1995).
- Willis, J. H. Effects of different levels of inbreeding on fitness components in *Mimulus guttatus*. *Evolution* **47**, 864–876 (1993).
- Fenster, C. B. & Ritland, K. Chloroplast DNA and isozyme diversity in two *Mimulus* species (Scrophulariaceae) with contrasting mating systems. *Am. J. Bot.* **79**, 1440–1447 (1992).
- Fu, Y.-B. & Ritland, K. Evidence for the partial dominance of viability genes contributing to inbreeding depression in *Mimulus guttatus*. *Genetics* **136**, 323–331 (1994).
- Latta, R. & Ritland, K. The relationship between inbreeding depression and prior inbreeding among populations of four *Mimulus* taxa. *Evolution* **48**, 806–817 (1994).
- Carr, D. E. & Fenster, C. B. Levels of genetic variation and covariation for *Mimulus* (Scrophulariaceae) floral traits. *Heredity* **72**, 606–618 (1994).
- Carr, D. E. & Dudash, M. R. Inbreeding depression in two species of *Mimulus* (Scrophulariaceae) with contrasting mating systems. *Am. J. Bot.* **83**, 586–593 (1996).
- Dudash, M. R. & Ritland, K. Multiple paternity and self-fertilization in relation to floral age in *Mimulus guttatus* (Scrophulariaceae). *Am. J. Bot.* **78**, 1746–1753 (1991).
- Comstock, R. E. & Robinson, H. F. The components of genetic variation in populations of biparental progenies and their use in estimating the average degree of dominance. *Biometrics* **2**, 254–266 (1948).
- Husband, B. C. & Schemske, D. W. Evolution of the magnitude and timing of inbreeding depression in plants. *Evolution* **50**, 54–70 (1996).
- Dudash, M. R., Carr, D. E. & Fenster, C. B. Five generations of enforced selfing and outcrossing in *Mimulus guttatus*: inbreeding depression variation at the population and family level. *Evolution* **51**, 54–65 (1997).
- Melchinger, A. E., Geiger, H. H. & Schnell, F. W. Epistasis in maize (*Zea mays* L.). 2. Genetic effects in crosses among early flint and dent inbred lines determined by three methods. *Theor. Appl. Genet.* **72**, 231–239 (1986).
- Pray, L. A. & Goodnight, C. J. Genetic variation in inbreeding depression in the red flour beetle *Tribolium castaneum*. *Evolution* **49**, 176–188 (1995).

Acknowledgements. We thank M. Lawrence and J. Crow for encouragement, J. Coors for statistical advice, and M. Bowers, L. Chao, D. Charlesworth, C. Cole, L. Delph, J. Dooley, C. Fenster and D. McCauley for comments on earlier versions of the manuscript. This research was supported by the NSF and Maryland Agricultural Experiment Station.

Correspondence and requests for materials should be addressed to M.R.D. (e-mail: md59@umail.umd.edu).

A nutrient-sensing pathway regulates leptin gene expression in muscle and fat

Jiali Wang, Rong Liu, Meredith Hawkins, Nir Barzilai & Luciano Rossetti

Diabetes Research and Training Center, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York 10461, USA

Leptin, the protein encoded by the *obese (ob)* gene, is synthesized and released in response to increased energy storage in adipose tissue^{1–4}. However, it is still not known how incoming energy is sensed and transduced into increased expression of the *ob* gene. The hexosamine biosynthetic pathway is a cellular 'sensor' of energy availability^{5–8} and mediates the effects of glucose on the expression of several gene products^{9–12}. Here we provide evidence for rapid activation of *ob* gene expression in skeletal muscle by glucosamine. Increased tissue concentrations of the end product of the hexosamine biosynthetic pathway, UDP-N-acetylglucosamine (UDP-GlcNAc), result in rapid and marked increases in leptin messenger RNA and protein levels (although these levels were much lower than those in fat). Plasma leptin levels and leptin mRNA and protein levels in adipose tissue also increase. Most important, stimulation of leptin synthesis is reproduced by either hyperglycaemia or hyperlipidaemia, which also increase tissue levels of UDP-N-acetylglucosamine in conscious rodents⁷. Finally, incubation of 3T3-L1 pre-adipocytes and L6 myocytes with glucosamine rapidly induces *ob* gene expression. Our findings

are the first evidence of inducible leptin expression in skeletal muscle and unveil an important biochemical link between increased availability of nutrients and leptin expression.

To determine whether the flux of carbons through the hexosamine biosynthetic pathway regulates *ob* expression in adipose tissue and skeletal muscle, we increased the tissue concentration of the end product of this pathway, UDP-*N*-acetylglucosamine (UDP-GlcNAc), by four different means (Fig. 1): first, glucosamine (GlcN) infusion⁶⁻⁸; second, uridine infusion^{7,8}; third, prolonged hyperglycaemia⁷; or fourth, prolonged hyperlipidaemia⁷. All experiments were carried out in conscious, unrestrained, chronically catheterized rats^{6,13} and included 3-h infusions of GlcN, uridine, lipid, glucose or vehicle in combination with euglycaemic or hyperglycaemic clamp studies. As described⁶⁻⁸, these four procedures all resulted in peripheral insulin resistance (decreased insulin-mediated glucose uptake) (Table 1).

Plasma leptin levels were increased (by two- to three fold; $P < 0.01$) in all experiments in which GlcN flux was stimulated as compared with the appropriate vehicle control study (Table 2). We measured leptin mRNA levels by reverse transcription with polymerase chain reaction (RT-PCR) (Fig. 2a) and northern blot analysis (Fig. 2b) in several tissues sampled at the completion of the studies. Although leptin mRNA was not detected in liver and lung (not shown), a distinct increase in leptin mRNA levels was induced by either hyperglycaemia or infusion of uridine, GlcN or lipid in both skeletal muscle (Fig. 2a, b) and adipose tissue (Fig. 3a, b). This stimulation of leptin gene expression was reversible.

Although the effect of GlcN on plasma leptin levels is probably due to its stimulation of leptin gene expression in adipose cells, the detection of leptin mRNA and protein in skeletal muscle was surprising and important. Comparison of the relative abundance of the leptin

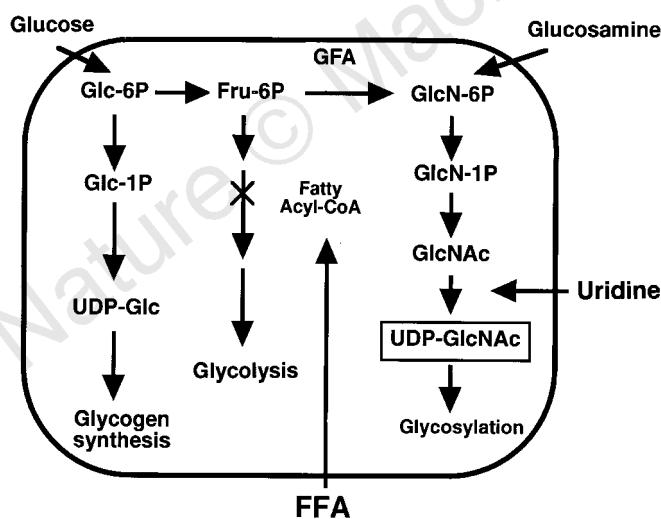


Figure 1 Sites of entry of glucose, GlcN and uridine into the hexosamine biosynthetic pathway in skeletal muscle and the role of free fatty acids (FFAs). Glucose is phosphorylated to produce glucose-6-phosphate (Glc-6P), which is used mainly in glycogen synthesis and glycolysis pathways. The hexosamine pathway receives ~1–3% of the incoming glucose through the conversion of fructose-6-phosphate (Fru-6P) to glucosamine-6-phosphate (GlcN-6P) by the rate-limiting enzyme glutamine:fructose-6-phosphate amidotransferase (GFA). Alternatively, infused glucosamine can enter skeletal muscle cells directly and is phosphorylated to produce GlcN-6P. Further metabolites are formed by subsequent acetylation and uridylation of GlcN-6P. Hence infused uridine, which contributes to the intracellular UDP/UTP pool, enhances the formation of downstream products of the hexosamine biosynthetic pathway. Increased FFA availability generates increased levels of fatty-acylCoA, leading to inhibition of glycolysis. This results in increased accumulation of Fru-6P and hence increased levels of substrate for GFA.

signals (mRNA and protein) in muscle and in adipose tissue showed that the striking induction of leptin expression in muscle was not due to contamination with adipose tissue. Furthermore, a transcript highly expressed in adipose tissue¹⁴, *AdipoQ*, was not detectable in muscle (Fig. 2c). Most important, skeletal muscle sampled at the completion of the GlcN infusions stained positively for leptin (Fig. 4c), whereas muscle sampled in the absence of GlcN (Fig. 4a) and 10 h after GlcN infusion (Fig. 4e) was immunonegative.

Leptin may regulate energy expenditure in skeletal muscle in part through the induction of expression of a new uncoupling protein, UCP-3 (ref. 15). However, in this case¹⁵ induction occurred following prolonged (four days of) exposure to leptin. Here, the skeletal muscle expression of UCP-3 was not affected by the short-term induction of leptin gene expression (Fig. 2d). The leptin mRNA

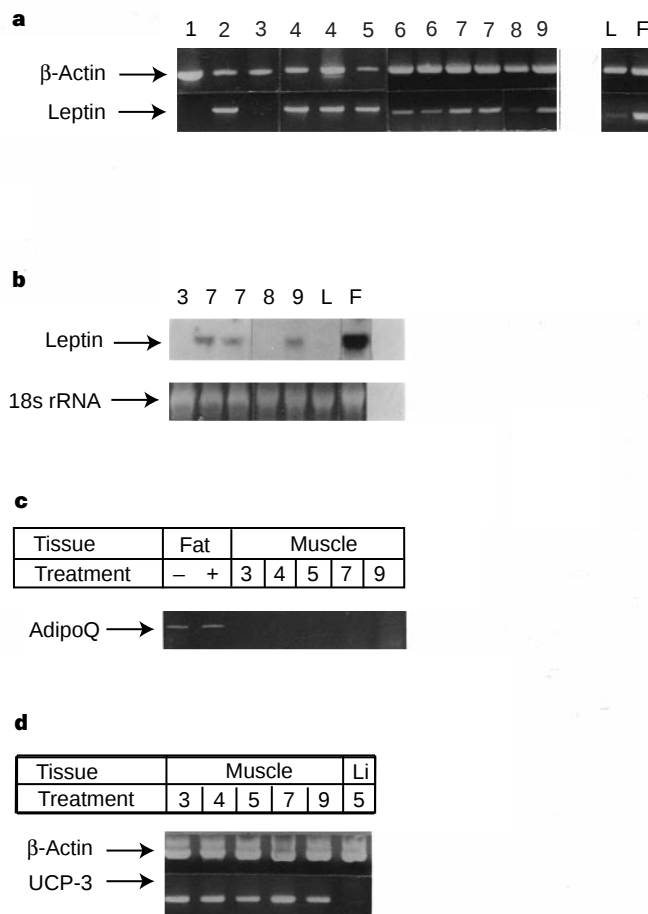


Figure 2 We studied levels of leptin mRNAs in different tissues by RT-PCR and northern blot analysis. **a**, Tissues were sampled at the completion of the *in vivo* studies. Lanes 1, 2, lean (L) and fat (F) as negative and positive controls, respectively, also shown at the right; lane 3, muscle sampled after 3-h infusions of vehicle (control); lanes 4, muscle sampled after 3-h infusions of uridine; lane 5, muscle sampled after 3 h of hyperglycaemia; lanes 6, muscle sampled 10 h after the completion of 3-h infusions of GlcN ('chase'); lanes 7, muscle sampled after 3-h infusion of GlcN ('pulse'); lane 8, muscle sampled 10 h after the completion of 3-h infusions of lipid ('chase'); lane 9, muscle sampled after 3-h infusions of lipid ('pulse'). **b**, Northern analysis of leptin mRNA in: skeletal muscle sampled after 3-h infusions of vehicle (control) (lane 3; 44 ± 3 arbitrary units (a.u.)); muscle sampled after 3-h infusions of GlcN (lanes 7; 919 ± 15 a.u.); muscle sampled 10 h after the completion of 3-h infusions of lipid (lane 8, 'chase'); muscle sampled after 3-h infusions of lipid (lane 9, 'pulse'); and muscle sampled from lean (L) and obese (F) Zucker rats. **c**, Expression of the *AdipoQ* gene was detected by RT-PCR using the same tissue sources as in **a**. **d**, mRNAs encoding UCP-3 were examined by RT-PCR. Experiments were replicated 3–5 times. Representative photographs are shown. Li, liver.

levels were also markedly increased in mixed white adipose tissue, brown adipose tissue, and, in particular, in subcutaneous white adipose tissue (Fig. 3a, b). In agreement with proposed links between increased energy availability, insulin resistance, and leptin expression in muscle, leptin mRNA was also readily detectable in skeletal muscle from Zucker fatty but not from Zucker lean rats (Fig. 2a, b). This finding was confirmed by immunohistochemistry (Fig. 4b, d).

We also measured leptin protein levels by radioimmunoassay. Adipose tissue and skeletal muscle levels of leptin were markedly increased (Table 2) in all experiments in which the tissue concentrations of UDP-GlcNAc were increased. Most important, the

Table 1 Plasma concentrations of glucose, insulin, free fatty acids and glucosamine, rate of glucose infusion and skeletal muscle concentrations of UDP-GlcNAc

Group	Hi GlcN	Hi Uridine	Hi Glc	Hi FFA	Con
Glucose (mM)	7.4 ± 0.2	7.5 ± 0.1	17.4 ± 0.2	7.4 ± 0.2	7.4 ± 0.2
Insulin (μU ml ⁻¹)	563 ± 39	549 ± 52	36 ± 9*	556 ± 52	561 ± 44
FFA (mM)	0.4 ± 0.2	0.4 ± 0.1	0.5 ± 0.2	1.2 ± 0.2*	0.3 ± 0.2
GlcN (mM)	0.9 ± 0.2*	<0.02	<0.02	<0.02	<0.02
GIR (mg kg ⁻¹ min ⁻¹)	26 ± 4*	28 ± 3*	29 ± 4*	25 ± 4*	39 ± 3
UDP-GlcNAc (ng g ⁻¹)	84 ± 13*	59 ± 5*	56 ± 6*	56 ± 4*	24 ± 3

These measurements were taken during the last 2 hours of hyperglycaemic clamp studies with low, physiological levels of insulin (Hi Glc), or during the last 2 hours of euglycaemic hyperinsulinaemic clamp studies with infusions of glucosamine (Hi GlcN), uridine (Hi Uridine), lipid emulsion (Hi free fatty acids, FFA) or saline (control, Con). GIR, rate of glucose infusion. *P < 0.01 versus control.

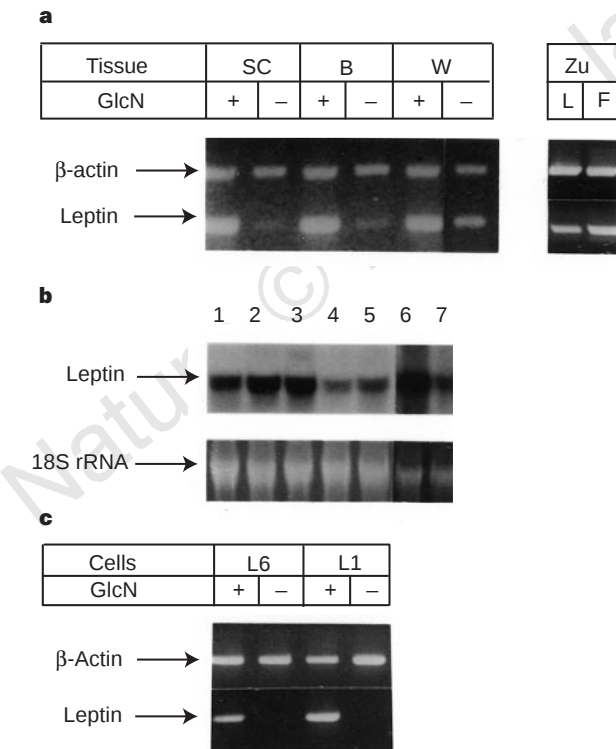


Figure 3 Leptin gene expression in fat tissues. **a**, RT-PCR analysis. Regulation of GlcN infusion in subcutaneous fat (SC), brown fat (B) and mixed white fat (W). L and F, mixed white fat from lean and obese Zucker rats, respectively. Experiments were replicated 3–5 times. A representative photograph is shown. Expression of β -actin was measured as a control. **b**, Northern blot analysis of leptin mRNA levels in either mixed white fat (lanes 1–3) or subcutaneous fat (lanes 4, 5), sampled after 3-h infusions of vehicle (control) (lanes 1, 4) or after 3-h infusions of glucosamine (lanes 2, 3, and 5); lanes 6 and 7 are from mixed white fat sampled from obese and lean Zucker rats. 18S rRNA was examined as a control. **c**, Induction of leptin gene expression by GlcN in cultured cells. L6, rat skeletal muscle cell L6; L1, 3T3 L1 preadipose cells. Experiments were replicated 3–5 times. A representative photograph is shown.

induction of leptin gene expression was independent of the procedure used to increase the accumulation of UDP-GlcNAc.

We next determined whether enhanced carbon flux through the hexosamine biosynthetic pathway could increase leptin mRNA and protein concentrations in cultured adipose and muscle cell lines. Short-term incubations of 3T3 preadipocytes and L6 myocytes in the presence of GlcN resulted in a marked and rapid induction of leptin mRNA expression (Fig. 3c).

Leptin may be vital in the sensing and regulation of lipid storage. In fact, leptin communicates the body's nutritional status to regulatory centres in the brain^{3,4,16}. In the long term, plasma leptin levels tend to increase in parallel with adipose stores and are strongly correlated with body fat mass¹⁷. However, the large variation in circulating leptin concentrations at similar levels of adiposity¹⁷ indicates that factors other than fat mass may be important in regulating serum leptin. The mechanism(s) for coupling metabolism and leptin expression has not been described. In the short term, regulation of leptin gene expression and circulating leptin levels has been observed with fasting^{18–20}, re-feeding^{2,17,20}, prolonged exposure to insulin and glucose^{21,22}, glucocorticoid administration²³, and

Table 2 Skeletal muscle, adipose tissue and plasma leptin levels resulting from increased concentrations of UDP-GlcNAc

Group	Hi GlcN	Hi uridine	Hi Glc	Hi FFA	Con
SM leptin (ng g ⁻¹)	2.9 ± 0.3*	1.9 ± 0.2*	1.7 ± 0.2*	1.9 ± 0.3*	ND (<0.3)
AT leptin (ng g ⁻¹)	23 ± 3*	21 ± 2*	19 ± 3*	21 ± 2*	15 ± 2
PL leptin (ng ml ⁻¹)	3.7 ± 0.2*	2.8 ± 0.3*	3.5 ± 0.3*	3.0 ± 0.2*	1.3 ± 0.1

These measurements were taken at the end of hyperglycaemic clamp studies with low, physiological levels of insulin (Hi Glc) or at the end of euglycaemic hyperinsulinaemic clamp studies with infusions of glucosamine (Hi GlcN), uridine (Hi Uridine), lipid emulsion (Hi free fatty acids, FFA) or saline (control, Con). AT, adipose tissue; ND, not detectable; PL, plasma; SM, skeletal muscle. *P < 0.01 versus control.

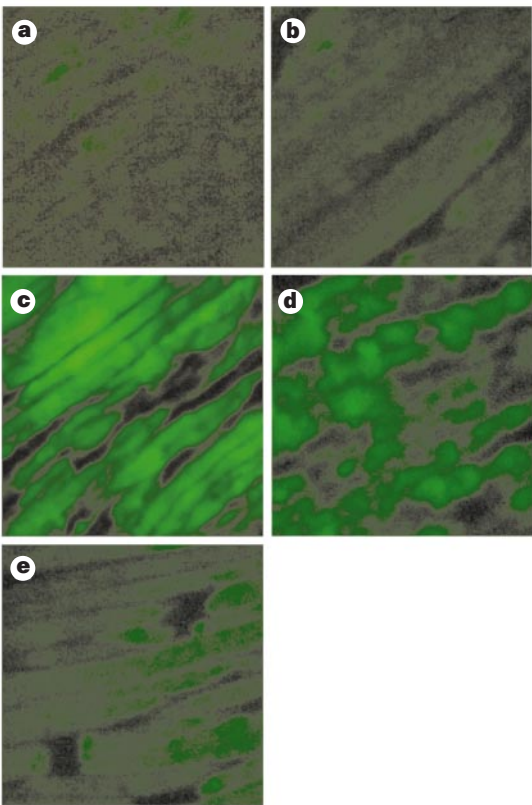


Figure 4 Immunohistochemical analysis of leptin in muscle tissue obtained from Sprague–Dawley rats. Results are from the completion of the control study (**a**) or of the GlcN infusion (**c**), or 10 h after the completion of infusions of GlcN (**e**). **b**, **d**, Muscle tissue obtained from lean (**b**) and obese (**d**).

increases in cellular cyclic AMP levels²³. The effect of insulin on leptin gene expression is unlikely to be direct²². Conversely, the decline in plasma leptin levels observed with prolonged fasting in humans can be completely reversed by an infusion of glucose designed to prevent the drop in both plasma glucose and insulin concentrations¹⁹. The rate of glucose uptake, which is rate limiting for triglyceride storage in adipose cells under most conditions, rather than insulin or glucose *per se*, may represent the signal for the regulation of leptin gene expression¹⁷. In contrast, in some studies^{17,24} leptin levels were, paradoxically, increased in the presence of insulin resistance. The association between insulin resistance and elevated leptin concentrations independent of adiposity and fat distribution raises the possibility that a common biochemical mechanism leads to both insulin resistance and leptin expression.

In adipose cells and skeletal muscle, glucose is transported and phosphorylated to produce glucose-6-phosphate. Glucose-6-phosphate is used mainly in two major pathways, namely, glycogen synthesis and glycolysis. However, ~1–3% of the incoming glucose is converted to fructose-6-phosphate and enters the hexosamine biosynthetic pathway (Fig. 1), whose main end product, UDP-GlcNAc, is the principal substrate that is used for O-linked and N-linked glycosylation of proteins. Glucose flux in primary cultures of adipose cells is proposed to be 'sensed' through the metabolism of hexose phosphates in this quantitatively minor pathway of intracellular glucose use. These *in vivo* observations have been reproduced *in vivo*^{6,25,26} and in isolated skeletal muscle²⁷. Most important, the accumulation of UDP-N-acetylhexosamines in skeletal muscle, an index of the flux through the hexosamine pathway, correlates highly with the degree of insulin resistance^{7,8}. We have shown previously that prolonged exposure to elevated levels of free fatty acids promotes increased flux of fructose-6-phosphate into the glucosamine pathway, even in the presence of stable or decreased glucose flux into skeletal muscle⁷, and that prolonged infusion of uridine results in both increased accumulation of UDP-GlcNAc and peripheral insulin resistance⁸. Thus this pathway is ideally positioned to link the rates of glucose and free fatty-acid flux to molecular and metabolic events.

The flux of glucose in the hexosamine biosynthetic pathway regulates the expression of several genes, such as the transforming growth factor- β ¹¹, transforming growth factor- α ¹⁰ and glutamine: fructose-6-phosphate amido transferase²⁸ genes. These effects of glucosamine on transcription are to be mediated, at least in part, through increased O-linked glycosylation of the ubiquitously expressed transcriptional factor Sp1 (ref. 12). O-linked glycosylation protects Sp1 from proteolytic degradation¹². We have shown here that this regulatory pathway also modulates the expression of *ob* in skeletal muscle and adipose tissue. Leptin is strongly expressed in white adipose tissue and at low levels in other tissues^{1,17}. Here we provide evidence of inducible leptin expression in skeletal muscle. This is consistent with the hypothesis that locally synthesized leptin may regulate energy expenditure¹⁵ and fat oxidation²⁹ in this tissue at least in part through a paracrine mechanism.

Our results support the idea that the hexosamine biosynthetic pathway functions as an energy-sensing device that is linked to the stimulation of leptin gene expression and secretion. Further studies will be needed to determine both the mechanism by which the activation of this biochemical pathway is transduced into increased transcription of the *ob* gene, and the relative contribution of this pathway to the overall regulation of leptin gene expression. □

Methods

Experimental procedures. We inserted catheters into the right internal jugular vein and the left carotid artery of male Sprague–Dawley and Zucker *fa/fa* and *fa/fa*– rats (Charles River Breeding Laboratories) as described^{6,13}. Studies were performed 5–7 days later in conscious rats fasted for ~6 h. Insulin clamp studies were performed as described^{6,13}. We designed four different experiments to generate increased flux into the glucosamine pathway.

In study 1 we infused GlcN into rats. GlcN ($15 \mu\text{mol kg}^{-1} \text{min}^{-1}$, $n = 6$), which enters the hexosamine pathway directly by being phosphorylated to produce GlcN-6-P, was infused for 3 h during euglycaemic hyperinsulinaemic ($18 \text{ mU kg}^{-1} \text{min}^{-1}$) clamp studies. To examine the reversibility of the effect of glucosamine infusion on leptin gene expression, we used another group of rats ($n = 3$) which received the same 3-h infusion, but we obtained tissue samples from these rats 10 h later.

In study 2, uridine was infused into rats. Euglycaemic hyperinsulinaemic ($18 \text{ mU kg}^{-1} \text{min}^{-1}$) clamp studies were performed in combination with infusions of uridine ($30 \mu\text{mol kg}^{-1} \text{min}^{-1}$, $n = 5$).

In study 3, hyperglycaemia (17 mM) and basal insulinaemia were maintained for 3 h using primed-continuous infusions of somatostatin (SRIF, $1.2 \mu\text{g kg}^{-1} \text{min}^{-1}$), variable glucose infusion, and insulin replacement ($0.7 \text{ mU kg}^{-1} \text{min}^{-1}$).

In study 4, continuous infusions of lipid emulsion (Liposyn 10%, Abbott Laboratories) were maintained for 3 h (high free fatty acids, 3 h, $n = 5$) at a rate of 1.5 ml h^{-1} during euglycaemic hyperinsulinaemic ($18 \text{ mU kg}^{-1} \text{min}^{-1}$) clamp studies. To examine the reversibility of the effect of lipid infusion on leptin gene expression, we used another group of rats ($n = 3$). These rats were infused with lipid emulsion for 3 h but tissue samples were obtained 10 h later.

Study 5 was a control. Vehicle time-control euglycaemic hyperinsulinaemic ($18 \text{ mU kg}^{-1} \text{min}^{-1}$) clamp studies (3 h) were performed in age- and weight-matched rats. Zucker rats were used as controls only.

Plasma samples for determination of insulin, leptin and GlcN concentrations were obtained at times 0, 60, 120 and 180 min. At the end of the *in vivo* studies, rats were anaesthetized (pentobarbital 60 mg per kg body weight, intravenously); the abdomen was quickly opened, and tissue samples were freeze-clamped *in situ* with aluminum tongs precooled in liquid nitrogen. Plasma glucose, insulin and GlcN and tissue UDP-GlcNAc concentrations were measured as described^{6–8}.

Cell culture. Skeletal muscle cells L6 (ATCC) and preadipose cells 3T3 L1 (a gift from C. S. Rubin) were re-seeded in Dulbecco's Modified Eagle's Medium (DMEM) containing 10% fetal bovine serum at a density of 10^6 per 100-mm dish, and were incubated overnight at 37°C in a 5% CO_2 incubator. Treatment was begun by incubation of the cells with insulin (7 mU) and dexamethasone ($1 \mu\text{M}$) for 1 h under the same conditions as above. GlcN was then added at a final concentration of 23 mM, and incubation continued for another 4 h. Cells were collected for RIA or RNA analysis.

Reverse transcription and amplification of complementary DNA. Total RNA from tissues and cultured cells was prepared following Clontech's protocol. The first-strand cDNA was synthesized from 3 μg total RNA in 20- μl final incubation volume by using the SuperScript preamplification system for first-strand cDNA synthesis (GIBCO) with random primers. PCR was carried out in a 50- μl reaction mixture containing 4 μl of the above first-strand cDNA, 5 μl of 10 $\times$ PCR buffer (Mg²⁺, Boehringer), 1 μl of 10-mM dNTP mix, 4 pmol of each primer and 2.5 U *Taq* DNA polymerase (GIBCO). The sequence of the upstream primer was 5'-TCC TAT CTG TCC TAT GTT CAA GCT GTG-3', and of the downstream primer was 5'-CAA CTG TTG AAG AAT GTC CTG CAG AGA-3'. The conditions for PCR were 94°C for 45 s, 69°C for 3.5 min (42 cycles) using GeneAmp PCR System 9600 (Perkin–Elmer). The expected RT-PCR product is 454 base pairs in length. To determine whether muscle samples were contaminated by fat tissues, oligonucleotides of sequence 5'-GTG TAC ATG AAA GAT GTG AAG GTG AGC-3' and 5'-ATA GCA TGG TCT ACT TCC AGA CAG TCA-3' were used as primers to amplify mRNA encoding AdipoQ, with a 60°C annealing temperature. The length of PCR product is 514 base pairs. Oligonucleotides used for amplification of mRNA encoding UCP-3 were 5'-GTT ACC TTT CCA CTG GAC ACA GCC AAG GTC-3' as the sense primer, and 5'-GAA GTT GTC AGT GAG CAG GTG GTA GTC CAG-3' as the antisense primer. The primers for β -actin were from Clontech.

Northern blot analysis. Total RNA (15 μg) from tissues were electrophoresed in formaldehyde-containing agarose gel, transferred to zeta-probe blotting membranes (Bio-Rad), and hybridized with the antisense cDNA probe. The rat leptin-coding region was obtained by RT-PCR and verified with DNA sequencing. The antisense probe was prepared by PCR with 0.1 μg cDNA, 4 pmol downstream primer, 5 μl of 10 $\times$ PCR buffer, 1 μl of 10 mM dGTP and dTTP, 18 μl of each [α -³²P]dCTP and [α -³²P]dATP (370 MBQ ml^{-1}) in 50 μl total reaction volume.

Immunofluorescence analysis. Muscle samples were embedded in OCT (Miles) and frozen in acetone-cooled dry ice. Cryosections (4 μ m) were air-dried for 20 min and washed three times in PBS for 15 min each time. Primary antibody (affinity-purified anti-murine leptin antibody, provided by M. Nicolson) was diluted (1:2000) in PBS containing 1% bovine serum albumin (BSA) and applied to each section for 1 h at room temperature. After washing with PBS three times for 15 min each time, the second antibody, rabbit immunoglobulin (fluorescein-linked whole antibody, Amersham), diluted (1:200) in PBS/1% BSA, was applied to the sections for 1 h. Sections were washed with PBS three times for 20 min each time and viewed on an immunofluorescence microscope (Olympus).

Quantification of leptin by radioimmunoassay. The homogenate from 0.5 g tissue in 1 ml lysis buffer (20 mM Imidazole/HCl, pH 6.8, 100 mM KCl, 1 mM $MgCl_2$, 10 mM EGTA, 10 mM NaF, 1 mM sodium molybdate, 1 mM sodium orthovanadate, 0.2% Triton X-100, 10 μ g ml⁻¹ aprotinin, 10 μ g ml⁻¹ leupeptin) was centrifuged at 3,000 r.p.m. at 4 °C in Sorvall RT6000B. The supernatant was transferred to a fresh tube for measurement of tissue leptin with the rat leptin RIA kit (Linco). Statistics were analysed with an unpaired Student's *t*-test and analysis of variance.

Received 19 January; accepted 7 April 1998.

- Zhang, Y. *et al.* Positional cloning of the mouse *obese* gene and its human homologue. *Nature* **372**, 425–432 (1994).
- Pelletier, M. A. *et al.* Effects of the obese gene product on body weight regulation in ob/ob mice. *Science* **269**, 540–543 (1995).
- Halaas, J. L. *et al.* Weight-reducing effects of the plasma protein encoded by the obese gene. *Science* **269**, 543–546 (1995).
- Campfield, L., Smith, F. J., Guisez, Y., Devos, R. & Burn, P. Recombinant mouse OB protein: evidence for a peripheral signal linking adiposity and central neural networks. *Science* **269**, 546–549 (1995).
- Marshall, S., Bacote, V. & Traxinger, R. R. Discovery of a metabolic pathway mediating desensitization of the glucose transport system: role of hexosamine biosynthesis in the induction of insulin resistance. *J. Biol. Chem.* **266**, 4706–4712 (1991).
- Rossetti, L., Hawkins, M., Chen, W., Gindi, J. & Barzilai, N. *In vivo* glucosamine infusion induces insulin resistance in normoglycemic but not in hyperglycemic conscious rats. *J. Clin. Invest.* **96**, 132–140 (1995).
- Hawkins, M., Barzilai, N., Hu, M., Chen, W. & Rossetti, L. Role of the glucosamine pathway in fat-induced insulin resistance. *J. Clin. Invest.* **99**, 2173–2182 (1997).
- Hawkins, M. *et al.* The tissue concentration of UDP-N-acetylglucosamine modulates the stimulatory effect of insulin on skeletal muscle glucose uptake. *J. Biol. Chem.* **272**, 4889–4895 (1997).
- Sayeski, P. P. & Kudlow, J. E. Glucose metabolism to glucosamine is necessary for glucose stimulation of transforming growth factor- α gene transcription. *J. Biol. Chem.* **271**, 15237–15243 (1996).
- Daniels, M. C. *et al.* Glucose regulation of transforming growth factor- α expression is mediated by products of the hexosamine biosynthesis pathway. *Mol. Endocrinol.* **7**, 1041–1048 (1993).
- McClain, D., Paterson, A., Roos, M., Wei, X. & Kudlow, J. Glucose and glucosamine regulate growth factor gene expression in vascular smooth muscle cells. *Proc. Natl Acad. Sci. USA* **89**, 8150–8154 (1992).
- Han, I. & Kudlow, J. E. Reduced O glycosylation of Sp1 is associated with increased proteasome susceptibility. *Mol. Cell Biol.* **17**, 2550–2558 (1997).
- Rossetti, L., Smith, D., Shulman, G. I., Papachristou, D. & DeFronzo, R. A. Correction of hyperglycemia with phlorizin normalizes tissue sensitivity to insulin in diabetic rats. *J. Clin. Invest.* **79**, 1510–1515 (1987).
- Hu, E., Liang, P. & Spiegelman, B. AdipoQ is a novel adipose-specific gene dysregulated in obesity. *J. Biol. Chem.* **271**, 10697–10703 (1996).
- Gong, D.-W., He, Y., Karas, M. & Reitman, M. Uncoupling protein-3 is a mediator of thermogenesis regulated by thyroid hormone, β -adrenergic agonists, and leptin. *J. Biol. Chem.* **272**, 24129–24132 (1997).
- Schwartz, M. W., Peskind, E., Raskind, M., Boyko, E. J. & Porte, D. J. Cerebrospinal fluid leptin levels: relationship to plasma levels and to adiposity in humans. *Nature Med.* **2**, 589–593 (1996).
- Flier, J. S. Leptin expression and action: new experimental paradigms. *Proc. Natl Acad. Sci. USA* **94**, 4242–4245 (1997).
- Ahima, R. S. *et al.* Role of leptin in the neuroendocrine response to fasting. *Nature* **382**, 250–252 (1996).
- Boden, G., Chen, X., Mozzolo, M. & Ryan, I. Effect of fasting on serum leptin in normal human subjects. *J. Clin. Endocrinol. Metab.* **81**, 3419–3423 (1996).
- Kolaczynski, J. W. *et al.* Responses of leptin to short-term fasting and refeeding in humans: a link with ketogenesis but not ketones themselves. *Diabetes* **45**, 1511–1515 (1996).
- Saladin, R. *et al.* Transient increase in obese gene expression after good intake or insulin administration. *Nature* **377**, 527–529 (1995).
- Kolaczynski, J. W. *et al.* Acute and chronic effects of insulin on leptin production in humans: studies *in vivo* and *in vitro*. *Diabetes* **45**, 699–701 (1996).
- MacDougald, O. A., Hwang, C. S., Fan, H. & Lane, M. D. Regulated expression of the obese gene product (leptin) in white adipose tissue and 3T3-L1 adipocytes. *Proc. Natl Acad. Sci. USA* **92**, 9034–9037 (1995).
- Segal, K. R., Landt, M. & Klein, S. Relationship between insulin sensitivity and plasma leptin concentration in lean and obese men. *Diabetes* **45**, 988–991 (1996).
- Baron, A. *et al.* Glucosamine induces insulin resistance *in vivo* by affecting GLUT4 translocation in skeletal muscle. *J. Clin. Invest.* **96**, 2792–2801 (1995).
- Hebert, L. *et al.* Overexpression of glutamine:fructose-6-phosphate amidotransferase in transgenic mice leads to insulin resistance. *J. Clin. Invest.* **98**, 930–936 (1996).
- Robinson, K. A., Sens, D. A. & Buse, M. G. Pre-exposure to glucosamine induces insulin resistance of glucose transport and glycogen synthesis in isolated rat skeletal muscles. *Diabetes* **42**, 1333–1346 (1993).
- Wang, J., Hawkins, M. & Rossetti, L. Glutamine:fructose-6-P amidotransferase (GFA) gene expression is induced by glucosamine flux in skeletal muscle. *Diabetes* **46** (suppl.), 325 (1997).
- Muoio, D. M., Dohn, G. L., Fiedorek, F. T., Tapscott, E. B. & Coleman, R. A. Leptin directly alters lipid partitioning in skeletal muscle. *Diabetes* **46**, 1360–1363 (1997).

Acknowledgements. We thank M. Hu and W. Chen for technical assistance. This work was supported by grants from the NIH and the American Diabetes Association and by the Albert Einstein Diabetes Research and Training Center. L.R. is the recipient of a Career Scientist Award from the Irma T. Hirsch Trust and N.B. of the Beeson Physician Faculty Scholar in Aging Award.

Correspondence and requests for materials should be addressed to L.R. (e-mail: rossetti@aecom.yu.edu).

Male-to-female sex reversal in *M33* mutant mice

Yuko Katoh-Fukui, Reiko Tsuchiya, Toshihiko Shiroishi*, Yoko Nakahara, Naoko Hashimoto, Kousei Noguchi† & Toru Higashinakagawa†

Mitsubishi Kasei Institute of Life Sciences, Machida, Tokyo 194-8511, Japan

* Mammalian Genetics Laboratory, National Institute of Genetics, Mishima, Shizuoka 411-8540, Japan

† Present address: Department of Biology, Waseda University, Shinjuku, Tokyo 169-8050, Japan

Polycomb genes in *Drosophila* maintain the repressed state of homeotic and other developmentally regulated genes^{1–4} by mediating changes in higher-order chromatin structure^{5–7}. *M33*, a mouse homologue of *Polycomb*, was isolated by means of the structural similarity of its chromodomain⁸. The fifth exon of *M33* contains a region of homology shared by *Drosophila* and *Xenopus*^{8,9}. In *Drosophila*, its deletion results in the loss of *Polycomb* function¹⁰. Here we have disrupted *M33* in mice by inserting a poly(A) capture-type *neo*^r targeting vector into its fifth exon. More than half of the resultant *M33*^{cterm}/*M33*^{cterm} mutant mice died before weaning, and survivors showed male-to-female sex reversal. Formation of genital ridges was retarded in both XX and XY *M33*^{cterm}/*M33*^{cterm} embryos. Gonadal growth defects appeared near the time of expression of the Y-chromosome-specific *Sry* gene¹¹, suggesting that *M33* deficiency may cause sex reversal by interfering with steps upstream of *Sry*. *M33*^{cterm}/*M33*^{cterm} mice may be a valuable model in which to test opposing views regarding sex determination.

Of 192 neomycin-resistant D3 embryonic stem (ES) cell clones, 17 were homologous recombinants. Seven independent clones were used to generate chimaeric males, four of which transmitted the

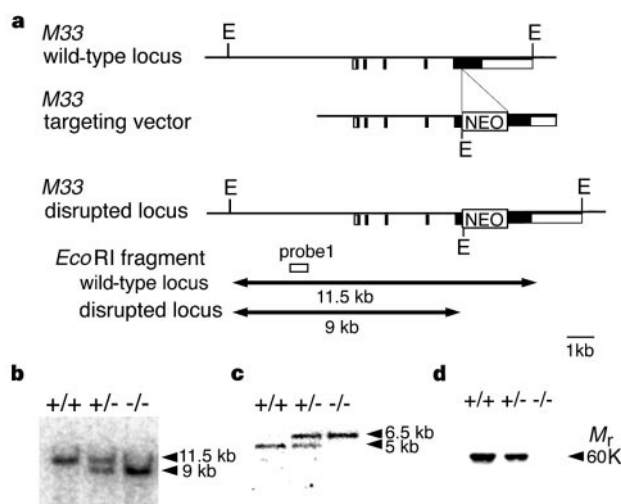


Figure 1 Targeted disruption of the *M33* gene. **a**, Genomic structure of *M33* wild-type locus, the targeting vector and the disrupted locus. Translated regions are shown as filled boxes. **b**, Southern blot hybridization of DNA of each genotype, digested with *Eco*RI and hybridized with probe 1 (shown in **a**). **c**, Northern blot hybridization of poly(A)⁺ RNA (0.5 μ g) from a 10.5 d.p.c. embryo of each genotype with *M33* cDNA as a probe. **d**, Detection of *M33* protein from 12.5 d.p.c. embryo of each genotype by western blot analysis.

disrupted allele (Fig. 1a) to the germ line. Two lines (A and B) were used for further analysis. $M33^{cterm}/+$ mice (*cterm* indicates disruption of the carboxy-terminal region of M33) grew normally without overt defects, and they were intercrossed to produce $M33^{cterm}/M33^{cterm}$ homozygotes (Fig. 1b). The ratios of $+/+$: $M33^{cterm}/+$: $M33^{cterm}/M33^{cterm}$ were 8:22:15 in embryos at 17.5 days *post coitum* (d.p.c.) and 28:38:17 in newborns. However, at 5–6 weeks *post partum*, only 30 of 317 progeny were $M33^{cterm}/M33^{cterm}$. Accordingly, 60% of $M33^{cterm}/M33^{cterm}$ mice died between the neonatal and the weaning periods. DNA analysis showed that all 17 that died between 8 and 30 days *post partum* (d.p.p.) were $M33^{cterm}/M33^{cterm}$. Mean body weight of these mice ($n = 11$) was 26% of that of their littermates. Of five examined, three had partial obstruction of the colon, and 8 of 17 had abnormalities of dental

occlusion (data not shown).

Northern blots of poly(A)⁺ RNA from $M33^{cterm}/M33^{cterm}$ embryos at 10.5 d.p.c. revealed a 6.5-kilobase (kb) transcript, 1.5 kb longer than expected for wild-type M33 messenger RNA (Fig. 1c). Transcription was presumably initiated from the wild-type M33 promoter and elongated through the fifth exon that included the inserted *neo*^r gene cassette, and translation was designed to terminate at the stop codon in the promoter of the *neo*^r gene cassette. Immunoblots of a lysate from 11.5 d.p.c. embryos probed with an antibody against the M33 C-terminal region showed no full-length protein (Fig. 1d). However, it is possible that the amino-terminal region of the M33 protein may remain in $M33^{cterm}/M33^{cterm}$ mice. Expression of such an N-terminal fragment could not be confirmed, because a specific antibody for this region is as yet unavailable.

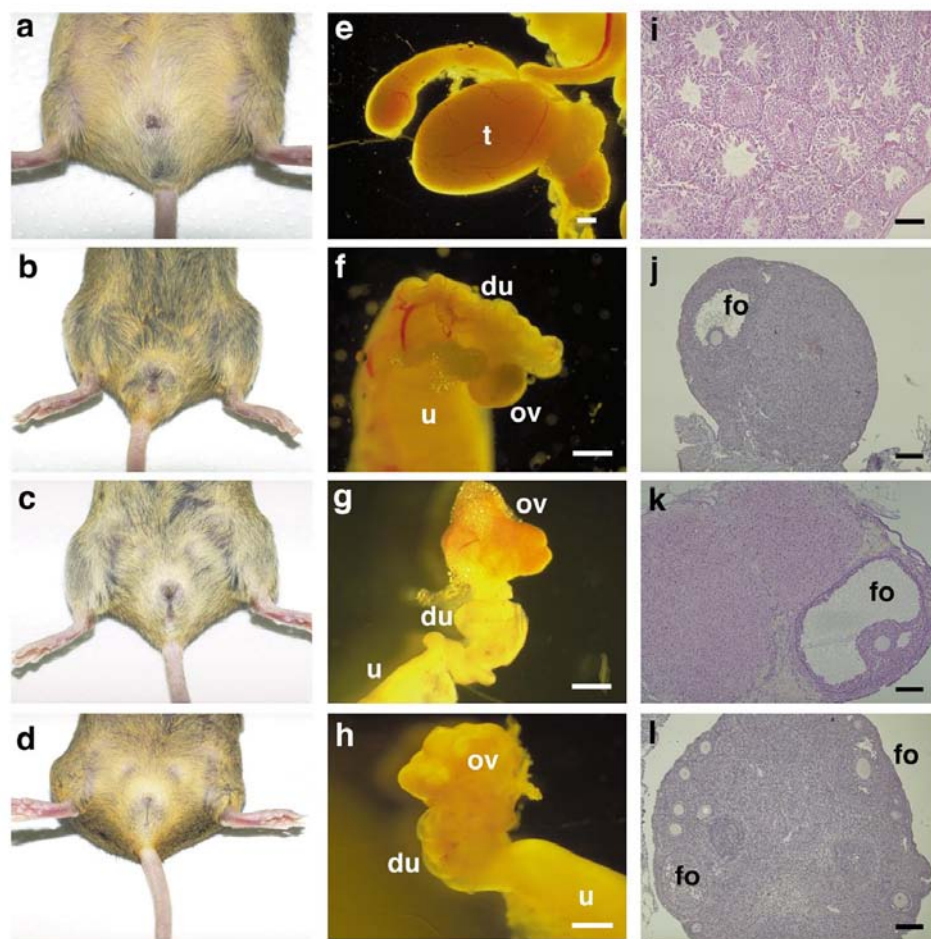


Figure 2 Male sex reversal in $M33^{cterm}/M33^{cterm}$ adult mice. **a–d**, External genitalia of adult mice at the same magnification. **a**, XY $M33 +/+$; **b**, XY $M33^{cterm}/M33^{cterm}$; **c**, XX $M33^{cterm}/M33^{cterm}$; **d**, XX $M33 +/+$. **e–h**, Gonads of adult mice. **e**, XY $M33 +/+$; **f**, XY $M33^{cterm}/M33^{cterm}$; **g**, XX $M33^{cterm}/M33^{cterm}$; **h**, XX $M33 +/+$. Scale bar, 1 mm. **i–l**, Cross-section of gonads of adult mice. **i**, XY $M33 +/+$; **j**, XY $M33^{cterm}/M33^{cterm}$; **k**, XX $M33^{cterm}/M33^{cterm}$; **l**, XX $M33 +/+$. Scale bar, 100 μ m. Abbreviations: du, oviduct; fo, follicle; ov, ovary; t, testis; u, uterus.

Table 1 Sex reversal of external genitalia and gonads in $M33^{cterm}/M33^{cterm}$ mice

	External genitalia			Gonads				
	male	female	intersex	testes	ovaries	testis + ovary	indistinct gonads	ovary + NE
$M33 +/+$	49*	0	0	38	0	0	0	0
$Sry + M33^{cterm}/+$	108†	0	0	79	0	0	0	0
$M33^{cterm}/M33^{cterm}$	2‡	4‡	6	0	6	3	3	0
$M33 +/+$	0	47*	0	0	29	0	0	0
$Sry - M33^{cterm}/+$	0	83§	0	0	61	0	0	0
$M33^{cterm}/M33^{cterm}$	0	12‡	4	0	14	0	0	2

Chromosomal sex was assessed by determining the presence or absence of the *Sry* gene by Southern blot hybridization using mouse *Sry* subclone p422.04 (ref. 12) as a probe. Sex of external genitalia and gonads were defined as follows: a mouse with a vaginal orifice was defined as female; a mouse with a penis was male; an intersex mouse had neither penis nor vaginal orifice; a gonad with at least one follicle was an ovary; a gonad with seminiferous tubules was a testis; and indistinct tissues adjacent to the oviduct and uterus exhibiting neither follicle nor seminiferous tubules were defined as indistinct gonads. NE indicates absence of gonad. Data were compiled from experiments using two independent ES cell clones. Eight XY $M33^{cterm}/M33^{cterm}$ mice containing all three hermaphrodite and nine XX $M33^{cterm}/M33^{cterm}$ mice were derived from one ES clone (line A), and remaining $M33^{cterm}/M33^{cterm}$ mice were derived from the other clone (line B).

* All mice tested were fertile ($n = 12$).

† All mice tested were fertile ($n = 14$).

‡ All mice were sterile.

§ All mice except two were fertile ($n = 24$).

Several $M33^{cterm}/M33^{cterm}$ adults had neither a vaginal orifice nor a penis, prompting us to investigate correlations between chromosomal sex, external genitalia and gonads. Chromosomal sex was determined by Southern blot analysis of tail DNA using the Y-chromosome-specific *Sry* gene as a probe¹². External genitalia of all XY $M33^{cterm}/+$ and XY $M33^{cterm}/M33^{cterm}$ mice were male type (Table 1; Fig. 2a). In contrast, those of the XY $M33^{cterm}/M33^{cterm}$ mutants were abnormal: four had female-type genitalia with clitoris and vaginal orifice (Fig. 2b); six had intersex-type genitalia with clitoris but without vaginal orifice; and two had male-type genitalia with a penis (Table 1). Most XX $M33^{cterm}/M33^{cterm}$ mice had normal female external genitalia (Fig. 2c), but some lacked vaginal orifice ($n = 4$ of 16). However, all XX $M33^{cterm}/M33^{cterm}$ mice were sterile during two months of continuous pairing with potent males, compared with all XX $M33^{cterm}/+$ mice and 92% of XX $M33^{cterm}/+$ mice which were fertile (Table 1). As expected, external genitalia of all XX $M33^{cterm}/+$ and XX $M33^{cterm}/+$ mice were female type (Table 1; Fig. 2d).

Internally, all XY $M33^{cterm}/+$ and XY $M33^{cterm}/+$ mice had bilateral testes, epididymides and vasa deferentia (Table 1; Fig. 2e, i). All XX $M33^{cterm}/+$ and XX $M33^{cterm}/+$ mice had paired ovaries, oviducts and uteri (Table 1; Fig. 2h, l). However, none of 12 XY $M33^{cterm}/M33^{cterm}$

mice had bilateral testes: six had paired uteri, oviducts and ovaries containing follicles (Fig. 2f, j); three had paired uteri, oviducts and indistinct gonads; and the remaining three had both testis and ovary (Table 1). In each hermaphrodite, the uterus and oviduct were located next to the ovary that contained follicles, and the epididymis and vas deferens were adjacent to the undescended testis.

We believe that these $M33^{cterm}/M33^{cterm}$ mice represent the first case in which male-to-female gonadal sex reversal results from a defect in a known recessive gene. XY females were created previously when the Y chromosome from *Mus musculus domesticus* was placed on the genetic background of laboratory mouse strains^{13,14}. Because the Y chromosome of $M33^{cterm}/M33^{cterm}$ mice was derived from the 129/Sv strain, which has a *Mus musculus molossinus* type Y chromosome, as does C57BL/6 (ref. 15), the male sex reversal we observed was not due to the presence of this divergent type of Y chromosome. The fact that all XY $M33^{cterm}/+$ and XY $M33^{cterm}/+$ mice were male is consistent with this conclusion.

Abnormalities of the reproductive system were also found in XX $M33^{cterm}/M33^{cterm}$ mutants. In 14 of 16 animals, ovaries on both sides were smaller than those of XX $M33^{cterm}/+$ and XX $M33^{cterm}/+$ mice (Table 1; Fig. 2g, k). These small ovaries contained follicles,

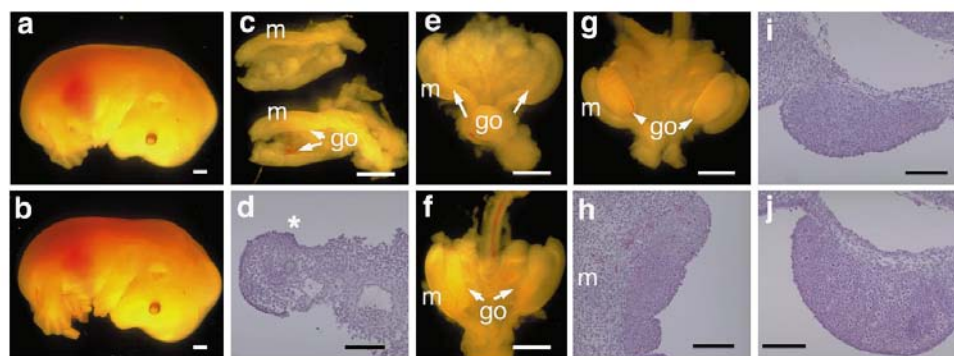


Figure 3 Developmental analysis of gonads. **a**, XY $M33^{cterm}/M33^{cterm}$ embryo at 13.5 d.p.c. **b**, XX wild-type embryo at 13.5 d.p.c. **c**, Urogenital ridges of $M33^{cterm}/M33^{cterm}$ (upper) and wild-type (bottom) embryos at 11.5 d.p.c. **d**, Transverse section of a $M33^{cterm}/M33^{cterm}$ gonad at 11.5 d.p.c.; asterisk indicates mesenchymal thickening. **e-g**, Gonads at 13.5 d.p.c. of: **e**, XY $M33^{cterm}/M33^{cterm}$; **f**, XX

wild-type; **g**, XY wild-type at 13.5 d.p.c. **h-j**, Sagittal section of gonads of: **h**, XY $M33^{cterm}/M33^{cterm}$ at 13.5 d.p.c.; **i**, XY $M33^{cterm}/M33^{cterm}$ at 15.5 d.p.c.; **j**, XX wild-type at 15.5 d.p.c. Scale bar: 100 μ m in **d**, **h**, **i** and **j**, otherwise 1 mm. Abbreviations: **go**, gonad; **m**, mesonephros.

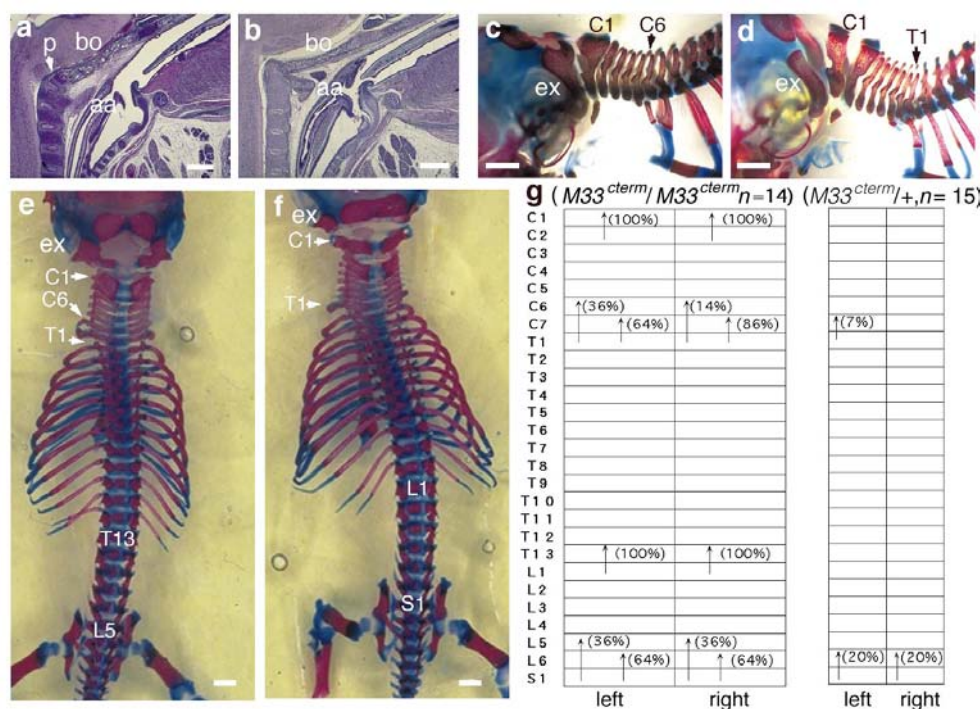


Figure 4 Skeletal analysis of $M33^{cterm}/M33^{cterm}$ newborn. **a**, **b**, Parasagittal sections: **a**, $M33^{cterm}/M33^{cterm}$; **b**, wild-type. **c**, **d**, Left side views of cervical and thoracic vertebrae: **c**, $M33^{cterm}/M33^{cterm}$; **d**, wild-type. **e**, **f**, Dorsal views of whole axial skeleton: **e**, $M33^{cterm}/M33^{cterm}$; **f**, wild-type; scapula, clavicle and forelimbs were removed. **g**, Frequencies of transformation. Ten $M33^{cterm}/M33^{cterm}$ mice were derived from line A, the rest from line B. Scale bars: 200 μ m in **a** and **b**; 1 mm in **c-f**. Abbreviations: **aa**, anterior arch of atlas; **bo**, basioccipital bone; **ex**, exoccipital bone; **p**, proatlas.

and in 12 animals they had a corpus luteum (Fig. 2k). Other developmental abnormalities in $XX\ M33^{cterm}/M33^{cterm}$ mutants included the absence of an ovary ($n = 2$ of 16).

$M33^{cterm}/M33^{cterm}$ embryos exhibited no overall retardation of development at any stage before birth (Fig. 3a, b), even though gonadogenesis was markedly delayed. Prominent gonadal defects appeared as early as 11.5 d.p.c. in both XY and XX embryos. At this time, urogenital ridges were morphologically indistinct ($n = 5$; Fig. 3c) and were distinguished only by mesenchymal thickening (Fig. 3d). By 13.5 d.p.c. these ridges were abnormally small (Fig. 3e–g). Indeed, they only developed to a size ($n = 7$) similar to that of 11.5 d.p.c. wild-type embryos (Fig. 3c, e). At 15.5 ($n = 3$) and 17.5 ($n = 4$), d.p.c. gonads were smaller than wild-type ovaries. No testis cord-like structures were present in XY embryos at 13.5 ($n = 4$; Fig. 3h), 15.5 ($n = 2$; Fig. 3i) or 17.5 d.p.c. ($n = 1$). Some gonads of 15.5 d.p.c. XY embryos had mitotic cells similar to those in wild-type females at the same stage (Fig. 3i, j). Morphological features of programmed cell death, which are prominent in *Ftz-F1* knockout mice¹⁶, were absent at 11.5 and 12.5 d.p.c. No size- or stage-specific defects in differentiation were evident in gonads of $M33^{cterm}/+$ embryos.

Point mutation of the human *SOX9* gene causes campomelic dysplasia (CMPD1) associated with male sex reversal^{17,18}. Because *M33* is located distal to mouse *Sox9* on chromosome 11 with a linkage of 5.3 cM (ref. 19) it is possible that *Sox9* was affected coincidentally during gene targeting in ES cells. Co-segregation with the disrupted *M33* gene may then have resulted in male sex reversal. To exclude this possibility we used recombination between *Sox9* and *M33* to generate $M33^{cterm}/M33^{cterm}$ mice in which ES cell (129/Sv)-derived *Sox9* was replaced by its counterpart from C57BL/6Njcl (B6). The resulting genotype was $Sox9^{B6}/M33^{cterm}/Sox9^{B6}/M33^{cterm}$. Two XY mice of this genotype both possessed a small ovary and an indistinct gonad that were associated with an oviduct and uterus. Thus male sex reversal in $M33^{cterm}/M33^{cterm}$ mice was not due to an affected *Sox9* gene, strongly supporting the idea that *M33* disruption was the proximate cause.

Coré *et al.* recently generated *M33* null knockout mice²⁰, with transformation of the axial skeleton as one of the manifested phenotypes. A sex-reversal phenotype was not recognized at that time because of a high incidence of postnatal lethality. However, re-examination of the specimens revealed that male-to-female sex reversal was indeed present (M. Djabali, personal communication). We also found skeletal defects in $M33^{cterm}/M33^{cterm}$ newborns (Fig. 4). The following transformations were noted in all mutants: first cervical vertebra (C1) to C2 of wild-type mice (Fig. 4c, d); C6 or C7 to first thoracic vertebra (T1) of wild-type mice (Fig. 4c, d); T13 to first lumbar vertebra (L1) of wild-type mice; and L5, or L6 to first sacral vertebra (S1) of wild-type mice (Fig. 4e, f).

Frequencies of skeletal transformation are summarized in Fig. 4g. Malformation of the exoccipital bone was present in some newborns ($n = 7$; Fig. 4c), probably as a consequence of fusion of the exoccipital bone with the manifested proatlas²¹ (Fig. 4a, b). Accordingly, the first vertebra posterior to the fused head bone was designated as the first cervical vertebra (C1) (Fig. 4c, d). This is consistent with previous results in which other mouse genes of the *Polycomb* group were targeted^{22–24}. Six vertebrosternal ribs were observed in all $M33^{cterm}/M33^{cterm}$ mutants, but scapula dysgenesis was absent. $M33^{cterm}/+$ mice showed posterior transformation, but the frequency was very low (Fig. 4g). Almost all *M33* null knockout mice with a mixed genetic background of BALB/C and 129/Ola died shortly after birth²⁰. Differences in survival between these mice and $M33^{cterm}/M33^{cterm}$ mutants may be due to either a difference in genetic background or a hypermorph of the $M33^{cterm}$ allele.

Sex determination seems to be a complex process involving both gonadal growth and genetic programmes for ovarian and testicular development, but their relative contributions remain unresolved. If rapid gonadal growth is essential for sex determination^{25,26},

growth retardation may be a cause of male-to-female sex reversal. Alternatively, determination of an ovary or a testis might be the result of competing programmes²⁷, with *Sry* acting to pre-empt the ovarian pathway. *M33* might regulate genes required for early gonadal development, and retardation of gonadogenesis in $M33^{cterm}/M33^{cterm}$ mice might result from misregulation of genes upstream of *Sry*. *M33* might also regulate repression of genes downstream of *Sry* in a manner analogous to *Polycomb* in *Drosophila*^{10,28}. Accordingly, the $M33^{cterm}/M33^{cterm}$ mutants provide an opportunity to look at components of the sex-determining pathway to find out how and where disruption of *M33* interferes with it. □

Methods

Targeting vector. *M33* genomic clones were isolated from a mouse 129/Sv genomic library (Stratagene) using *M33* complementary DNA amplified by the polymerase chain reaction (PCR) as a probe. A targeting vector, PCTG-5, was constructed by inserting a 1.3-kb *EcoRI*–*BamHI* fragment of pKJ2 (ref. 29), containing a Pkg-1 promoter and a *neo^r* gene without a polyadenylation signal, into the *BamHI* restriction site of the fifth exon of *M33*. This construct was expected to delete the C-terminal 361 amino acids from the 519 residues of the intact *M33* protein.

ES cells and *M33* mutant mice. D3 ES cells were maintained as described³⁰. Cells (4×10^7) were electroporated using a Gene Pulser (Bio Rad) at 220 V, 500 μ F in phosphate buffered saline (PBS) containing 20 μ g of PCTG-5. Male chimaeras were bred with C57BL/6Njcl (B6) females. $M33^{cterm}/+$ F₁ males were mated with B6 females to expand the pool of F₂ mice carrying the mutant allele. $M33^{cterm}/M33^{cterm}$ mutants were obtained by intercrossing these F₂ mice.

Western blotting. Protein (20 μ g) from a 12.5 d.p.c. embryo of each genotype was immunoblotted with affinity-purified polyclonal antiserum raised against the C-terminal 229 amino acids encompassing the conserved region of the *M33* protein.

Test matings and gonadal analysis. Test matings were done 8–16 weeks after birth. Surgery was performed to inspect the gonads of adult mice at 17–46 weeks after birth and those of embryos at 11.5 d.p.c. to 17.5 d.p.c. $M33^{cterm}/M33^{cterm}$ mice with male-type external genitalia were paired with B6 females; those with female-type external genitalia were paired with B6 males. No test matings were done on $M33^{cterm}/M33^{cterm}$ mutants with intersex type external genitalia.

Histology. Tissues were fixed on Bouin's solution, embedded in paraffin, and sectioned at a thickness of 5 μ m. Sections were stained with haematoxylin and eosin, dehydrated and mounted with coverslips for photomicrography.

***Sox^{B6}/M33^{cterm}/Sox^{B6}/M33^{cterm}* recombinants.** Mice heterozygous for the $M33^{cterm}$ mutant allele, offspring of crosses between male chimaeras and B6, were backcrossed with B6 males. Progeny were typed for microsatellite markers *D11Mit167*, located in the interval between *Sox9* and *M33*, and *D11Mit58*, located ~ 6.5 cM proximal to *Sox9*, based on the simple sequence length polymorphism between 129/Sv and B6. Heterozygotes were intercrossed, and mice homozygous for *D11Mit58^{B6}-Sox^{B6}-D11Mit167^{B6}-M33^{cterm}* were generated.

Received 23 September 1997; accepted 8 April 1998.

- Lewis, E. B. A gene complex controlling segmentation in *Drosophila*. *Nature* **276**, 565–570 (1978).
- Wede, C., Harding, K. & Levine, M. Spatial regulation of Antennapedia and Bithorax gene expression by the *Polycomb* locus in *Drosophila*. *Cell* **44**, 739–748 (1986).
- Beachy, P. A., Helfand, S. L. & Hogness, D. S. Segmental distribution of bithorax complex proteins during *Drosophila* development. *Nature* **313**, 545–551 (1985).
- Kuziora, M. A. & McGinnis, W. Different transcripts of the *Drosophila Abd-B* gene correlate with distinct genetic sub-functions. *EMBO J.* **7**, 3233–3244 (1988).
- Paro, R. & Hogness, D. S. The *Polycomb* protein shares a homologous domain with a heterochromatin-associated protein of *Drosophila*. *Proc. Natl Acad. Sci. USA* **88**, 263–267 (1991).
- Zink, B. & Paro, R. *In vivo* binding pattern of a trans-regulator of homeotic genes in *Drosophila melanogaster*. *Nature* **337**, 468–471 (1989).
- Orlando, V. & Paro, R. Mapping *Polycomb*-repressed domains in the Bithorax complex using *in vivo* formaldehyde cross-linked chromatin. *Cell* **75**, 1187–1198 (1993).
- Pearce, J. J. H., Singh, P. B. & Gaunt, S. J. The mouse has a *Polycomb*-like chromobox gene. *Development* **114**, 921–929 (1992).
- Reijnen, M. J. *et al.* *Polycomb* and *bmi-1* homologs are expressed in overlapping patterns in *Xenopus* embryos and are able to interact with each other. *Mech. Dev.* **53**, 35–46 (1995).
- Müller, J. Transcriptional silencing by the *Polycomb* protein in *Drosophila* embryos. *EMBO J.* **14**, 1209–1220 (1995).
- Hacker, A., Capel, B., Goodfellow, P. & Lovell-Badge, R. Expression of *Sry*, the mouse sex determining gene. *Development* **121**, 1603–1614 (1995).

12. Gubbay, J. *et al.* A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. *Nature* **346**, 245–250 (1990).
13. Eichler, E. M., Washburn, L. L., Whitney, J. B. III & Morrow, K. E. *Mus poschiavinus* Y chromosome in the C57BL/6J murine genome causes sex reversal. *Science* **217**, 535–537 (1982).
14. Eichler, E. M. & Washburn, L. L. Inherited sex reversal in mice: identification of a new primary sex-determining gene. *J. Exp. Zool.* **228**, 297–304 (1983).
15. Nagamine, C. M. *et al.* The musculus-type Y chromosome of the laboratory mouse is of Asian origin. *Mamm. Genome* **3**, 84–91 (1992).
16. Luo, X., Ikeda, Y. & Parker, K. L. A cell-specific nuclear receptor is essential for adrenal and gonadal development and sexual differentiation. *Cell* **77**, 481–490 (1994).
17. Wagner, T. *et al.* Autosomal sex reversal and campomelic dysplasia are caused by mutations in and around the SRY-related gene SOX9. *Cell* **79**, 1111–1120 (1994).
18. Foster, J. W. *et al.* Campomelic dysplasia and autosomal sex reversal caused by mutations in an SRY-related gene. *Nature* **372**, 525–530 (1994).
19. Uchida, K. *et al.* Exclusion of *Sox9* as a candidate for the mouse mutant Tail-short. *Mamm. Genome* **7**, 481–485 (1996).
20. Coré, N. *et al.* Altered cellular proliferation and mesoderm patterning in Polycomb-M33-deficient mice. *Development* **124**, 721–729 (1997).
21. Kessel, M., Balling, R. & Gruss, P. Variation of cervical vertebrae after expression of a *Hox-1.1* transgene in mice. *Cell* **61**, 301–308 (1990).
22. Takiyama, Y. *et al.* Targeted disruption of the mouse homologue of the *Drosophila* polycomb gene lead to altered anteroposterior patterning and neural crest defects. *Development* **124**, 3673–3682 (1997).
23. van der Lugt, N. M. T. *et al.* Posterior transformation, neurological abnormalities, and severe hematopoietic defects in mice with a targeted deletion of the *bmi-1* proto-oncogene. *Genes Dev.* **8**, 757–769 (1994).
24. Akasaka, T. *et al.* A role for *mel-18*, a Polycomb group-related vertebrate gene, during the anteroposterior specification of the axial skeleton. *Development* **122**, 1513–1522 (1996).
25. Mittwoch, U. Sex differentiation in mammals and tempo of growth probabilities vs. switches. *J. Theor. Biol.* **137**, 445–455 (1989).
26. Cattanchi, B. M. Sex-reversed mice and sex determination. *Ann. N. Y. Acad. Sci.* **513**, 27–39 (1987).
27. Burgoyne, P. S. & Palmer, S. J. The genetics of XY sex reversal in the mouse and other mammals. *Semin. Dev. Biol.* **2**, 277–284 (1991).
28. Bunker, C. A. & Kingston, R. E. Transcription repression by *Drosophila* and Mammalian Polycomb group proteins in transfected mammalian cells. *Mol. Cell. Biol.* **14**, 1721–1732 (1994).
29. Boer, P. H. *et al.* Polymorphisms in the coding and noncoding regions of murine *Pgk-1* alleles. *Biochem. Genet.* **28**, 299–308 (1990).
30. Robertson, E. J. in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach* (ed. Robertson, E. J.) 71–112 (IRL, Oxford, 1987).

Acknowledgements. We thank T. Takeuchi, Y. Takiyama, H. Koseki and T. Akasaka for advice and discussion; M. Djabali for communicating unpublished results; and R. A. Shiurba for critical reading of the manuscript.

Correspondence and requests for materials should be addressed to Y.K.-F. (e-mail: kan@libra.la.m.kagaku.co.jp).

Premotor commands encode monocular eye movements

Wu Zhou & W. M. King

University of Mississippi Medical Center, Departments of Neurology and Anatomy, 2500 North State Street, Jackson, Mississippi 39216, USA

Binocular coordination of eye movements is essential for stereopsis (depth perception) and to prevent double vision. More than a century ago, Hering and Helmholtz debated the neural basis of binocular coordination. Helmholtz¹ believed that each eye is controlled independently and that binocular coordination is learned. Hering² believed that both eyes are innervated by common command signals that yoke the eye movements (Hering's law of equal innervation). Here we provide evidence that Hering's law is unlikely to be correct. We show that premotor neurons in the paramedian pontine reticular formation that were thought to encode conjugate^{3–6} velocity commands for saccades (rapid eye movements) actually encode monocular commands for either right or left eye saccades. However, 66% of the abducens motor neurons, which innervate the ipsilateral lateral rectus muscle, fire as a result of movements of either eye. The distribution of sensitivity to ipsilateral and contralateral eye movements across the abducens motor neuron pool may provide a basis for learning binocular coordination in infancy and adapting it throughout life.

Abducens motor neurons are the final common pathway along which neural commands travel to innervate the lateral rectus muscle of the eye unilaterally. Figure 1 illustrates the discharge pattern of an abducens motor neuron located on the right side during 'smooth pursuit' movements in which only one eye moved; we name these movements 'monocular pursuit' movements. Firing rate decreased

when the ipsilateral eye adducted and increased when it abducted (Fig. 1a, thin trace). This is the expected discharge pattern of an abducens motor neuron⁷. Unexpectedly, however, during monocular pursuit movements of the contralateral eye (Fig. 1b, thick trace), the cell's firing rate was also modulated even though the ipsilateral eye was stationary. To examine this relationship quantitatively, we calculated mean eye position and firing rate over 50-ms intervals (Fig. 1c, d). Firing rate is linearly related to ipsilateral eye position (Fig. 1c, black circles) and contralateral eye position (Fig. 1d, open triangles) eye position. The regression coefficient relating firing rate to ipsilateral eye position ($3.3 \text{ spikes s}^{-1} \text{ deg}^{-1}$) is greater than the coefficient relating firing rate to contralateral eye position ($2.3 \text{ spikes s}^{-1} \text{ deg}^{-1}$). Not every abducens motor neuron exhibits binocular discharge characteristics. Figure 2 illustrates the activity of a motor neuron with a monocular discharge pattern: its firing rate is related to ipsilateral eye position (Fig. 2a, c, $4.1 \text{ spikes s}^{-1} \text{ deg}^{-1}$, $P < 0.001$) but not to contralateral eye position (Fig. 2b, d, $0.002 \text{ spikes s}^{-1} \text{ deg}^{-1}$, $P > 0.97$).

To determine the distribution of sensitivity to ipsilateral and contralateral eye movements across the motor neuron pool, we obtained single unit recordings from 136 motor neuron axons. Multiple regression analyses showed that many motor neurons were monocular, with regression coefficients relating firing

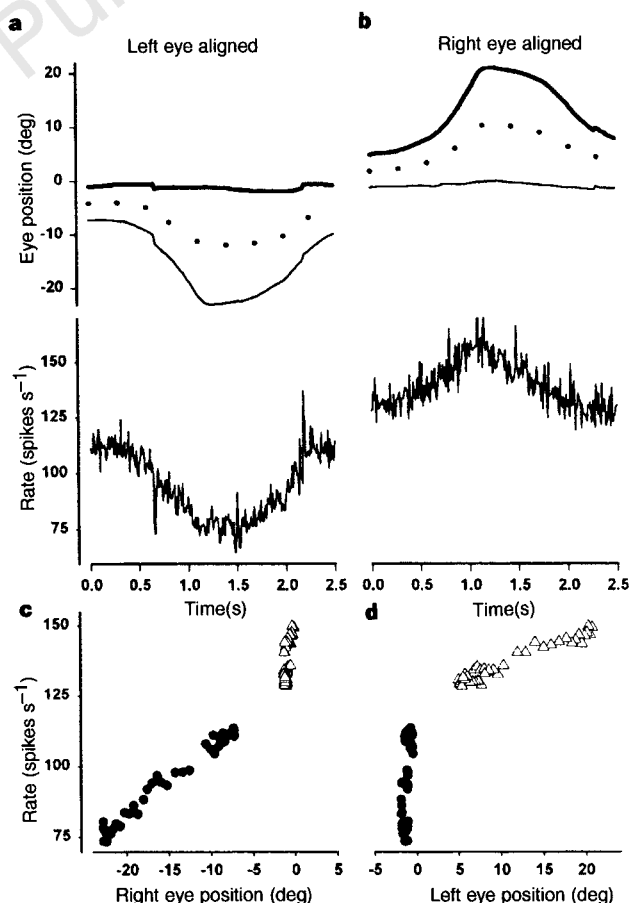


Figure 1 Discharge pattern of a binocular abducens motor neuron recorded in the VIth nerve. **a**, Monocular pursuit, aligned with the monkey's left eye, of a visible target that is moved towards or away from the animal in a sagittal plane. Upper panel, the left eye (thick trace) is relatively stationary while the right eye (thin trace) tracks the target. Eye movement to the left is indicated by downward deflection of the traces, and movement to the right is indicated by an upward trace deflection (in Figs 1, 2 and 4). The dotted trace indicates the conjugate eye position. Lower panel, rate of firing of abducens motor neurons. **b**, Monocular pursuit aligned with the right eye. **c**, **d**, Firing rate correlates linearly with right eye position (**c**, black circles) and left eye position (**d**, open triangles).

rate either to ipsilateral (Fig. 3a, black circles) or to contralateral (Fig. 3a, black triangles) eye position ($P < 0.05$). The mean sensitivity of monocular motor neurons to eye position was $3.84 \pm 1.03 \text{ spikes s}^{-1} \text{ deg}^{-1}$ ($N = 37$) for ipsilateral eye units and $2.72 \pm 1.48 \text{ spikes s}^{-1} \text{ deg}^{-1}$ ($N = 9$) for contralateral eye units. We only classified cells as monocular if one of the regression coefficients was not significantly different from zero ($P > 0.05$, 32 out of 46 cells) or if the regression, R^2 , was unchanged by including the smaller coefficient ($P > 0.05$, 14 out of 46 cells). The remaining 90 motor neurons were binocular, with significant regression coefficients for the ipsilateral ($3.39 \pm 1.19 \text{ spikes s}^{-1} \text{ deg}^{-1}$, grey circles) and contralateral ($1.40 \pm 0.63 \text{ spikes s}^{-1} \text{ deg}^{-1}$, grey triangles) eye.

The distribution of regression coefficients is not uniform (Fig. 3a, upper panel). To illustrate this effect, we calculated a numerical index called ocular selectivity (Fig. 3a, lower panel). Monocular motor neurons have ocular selectivities clustered at either end of the distribution (Fig. 3a, lower panel, black bars). Of these monocular motor neurons, 27% are monocular with respect to the ipsilateral eye (mean ocular selectivity = 0.86, s.d. = 0.09) and 7% are monocular with respect to the contralateral eye (mean = -0.94, s.d. = 0.06). Ocular selectivities of binocular motor neurons are distributed in between (66%, grey bars), with a clear preference for the ipsilateral eye ($0 < \text{ocular selectivity} < 1$, mean = 0.34, s.d. = 0.31).

As left or right eye position is a linear sum of conjugate and vergence eye positions, these data do not resolve the conflict between Hering's and Helmholtz's hypotheses. For example, cells

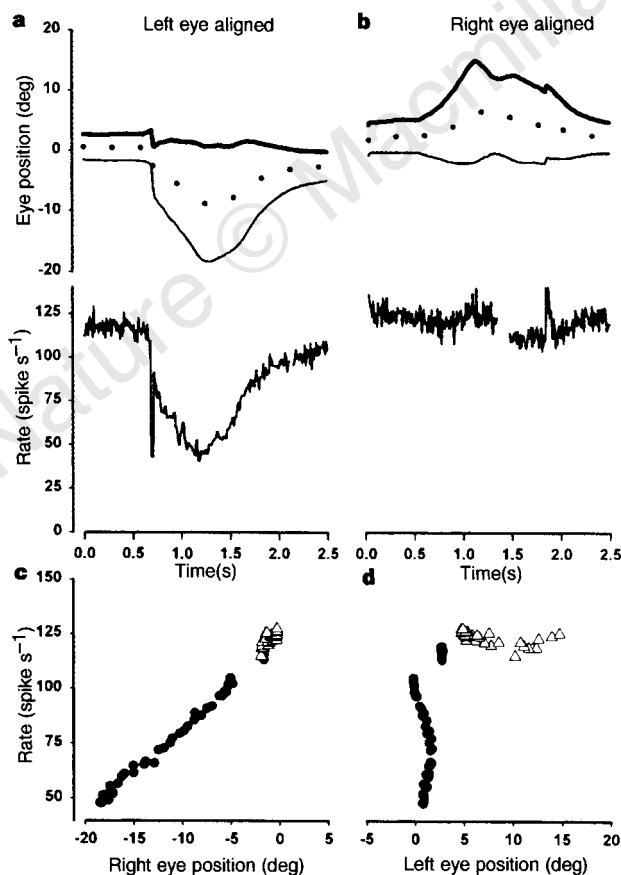


Figure 2 Discharge pattern of a monocular abducens motor neuron. **a**, Monocular pursuit with the ipsilateral eye (right eye, thin trace). **b**, Monocular pursuit with the left eye (thick trace). **c**, **d**, Firing rate correlates linearly with right eye position (**c**) but not with left eye position (**d**). Black circles represent data from trials in **a** (left eye aligned), and open triangles represent data from trials in **b** (right eye aligned).

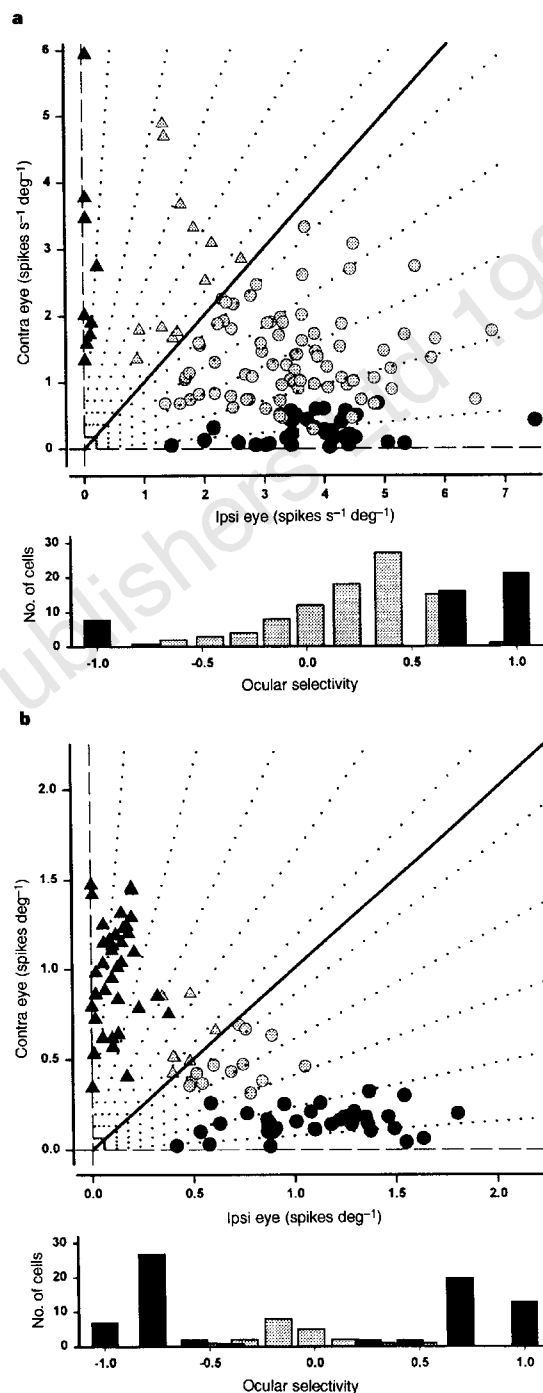


Figure 3 The distribution of regression coefficients relating neuronal discharge to contralateral or ipsilateral eye position or saccade amplitude. **a**, Upper panel, regression coefficients for abducens motor neurons. Black symbols represent monocular motor neurons; grey symbols represent binocular units. Circles represent cells with ipsilateral (ipsi) eye preference, and triangles represent cells with contralateral (contra) eye preference. Lower panel, distribution of ocular selectivity. Black bars correspond to monocular motor neurons, and grey bars correspond to binocular motor neurons. **b**, Regression coefficients (upper panel) and ocular selectivity (lower panel) for EBNS. Symbols as in **a**. Note the expanded scale in the upper panel. In each upper panel, the dotted lines define 9-degree wedges centred about the axes and about the unity slope line (thick black line). These wedges define the bins used in the ocular selectivity plots. Ocular selectivity for motor neurons is defined as $(K_{\text{ipsi}} - K_{\text{contra}})/(K_{\text{contra}} + K_{\text{ipsi}})$, where K_{contra} and K_{ipsi} are the regression coefficients relating firing rate to contralateral or ipsilateral eye position, respectively. For EBNS, ocular selectivity is defined similarly except the regression coefficients relate the number of spikes for saccade amplitude. Ocular selectivity is >0 for ipsilateral eye preference.

that we have classified as monocular might simply receive balanced conjugate and vergence input signals that cancel one another out during monocular pursuit⁸. To distinguish between these hypotheses one can record the activity of premotor cells during disjunctive eye movements that strongly dissociate right and left eye position. Excitatory burst neurons (EBNs) in the paramedian pontine reticular formation (PPRF) project directly to motor neurons⁹. During saccades, EBNs discharge a burst of spikes that encodes eye velocity. The burst is integrated by a neuronal circuit involving nucleus prepositus hypoglossi (NPH) neurons to produce an eye-position command⁹. Thus the number of spikes in the burst is correlated with the amplitude of the saccade¹⁰. For example, during ipsilateral conjugate eye movements of 10 degrees, the EBN shown in Fig. 4a emitted bursts of 8–11 spikes. During monocular 10-degree saccades executed by the right eye (Fig. 4b), the EBN continued to emit bursts of 8–10 spikes despite the reduction of conjugate saccade amplitude (Fig. 4b, grey traces) to 5 degrees in these trials. This cell never discharged during conjugate saccades to the contralateral side. However, during asymmetric vergence saccades with contralateral conjugate components (Fig. 4c), the cell emitted bursts of 4–7 spikes, consistent with the ipsilateral saccade of 3–5 degrees made by the right eye. These data indicate that this cell monocularly encodes movements of the right eye. Indeed, multiple regression analysis of data from many trials shows that the number of spikes in a burst encodes the amplitude of right eye saccades (Fig. 5a;

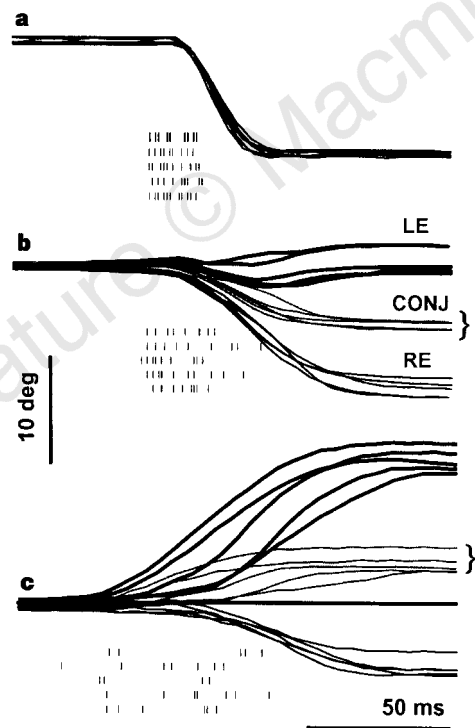


Figure 4 Excitatory burst neurons in the left PPRF encode monocular eye movements. **a**, Leftward conjugate saccades of 10 degrees. **b**, Monocular saccades of 10 degrees in the right eye. **c**, Convergent saccades of about 10 degrees. The thick traces indicate left eye (LE) position; the bracketed traces indicate conjugate position (CONJ); and the thin traces indicate right eye (RE) position. Each record consists of five superimposed trials. Neural action potentials are represented as vertical tick marks aligned with the eye movement records.

-1.00 ± 0.05 spikes deg^{-1} , $P < 0.001$) but not the amplitude of left eye saccades (Fig. 5b; 0.02 ± 0.04 spikes deg^{-1} , $P > 0.55$).

In Fig. 5c, d, we show similar data from another EBN but include divergence as well as convergence trials. Data from divergence trials (black triangles) overlay data from convergence (grey inverted triangles) and conjugate (grey circles) saccade trials when plotted against the amplitude of the left eye saccade. This EBN encoded the amplitude of the left eye saccade (Fig. 5c; 1.29 ± 0.04 spikes deg^{-1} , $P < 0.001$) but not of the right eye saccade (Fig. 5d; -0.06 ± 0.04 spikes deg^{-1} , $P > 0.11$). In Fig. 3b we show that most of the 96 studied EBNs were monocular (79%, black symbols). Only 5.3% (five cells) clustered near the unity slope line could be said to encode conjugate saccade amplitude.

We further tested each of the EBNs that we classified as monocular for any possible relationship to conjugate saccade amplitude. We selected a subset of trials for which the amplitude of the saccades made by the eye associated with the larger regression coefficient was constant, while the other eye executed saccades of varying amplitude (thus conjugate amplitude varied over a large range). For each neuron, we calculated the regression of the number of spikes related to conjugate saccade amplitude; in every case, the regression was not significant ($P > 0.05$). Furthermore, for every monocular neuron, the correlation coefficient, R , of the monocular regression (amplitude of the ipsilateral or contralateral eye saccade) was significantly greater than the regression for conjugate amplitude ($P < 0.05$).

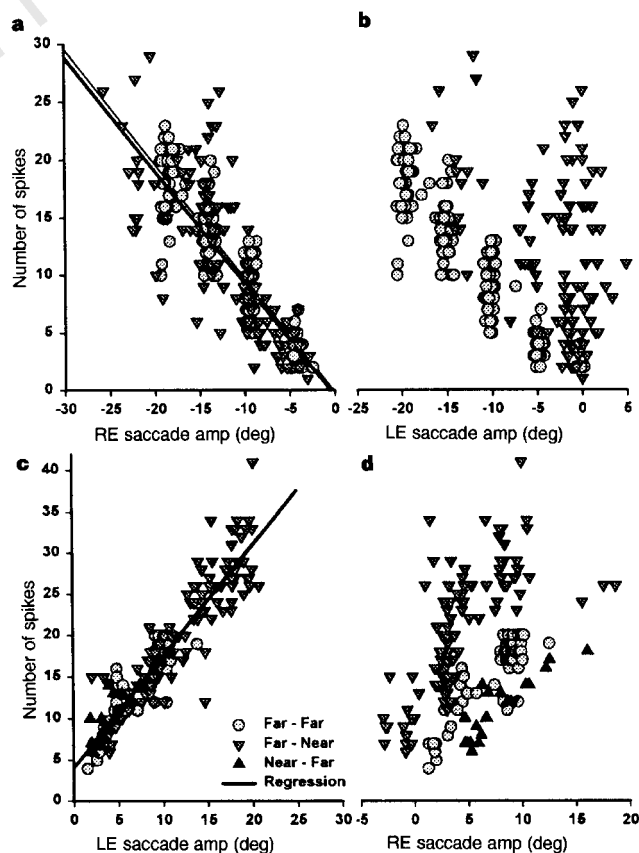


Figure 5 Examples of monocular discharge patterns of 2 EBNs. The number of spikes emitted is proportional to the amplitude of either the right eye saccade (**a**) but not of the left eye saccade (**b**) during disjunctive eye movements. For the other cell, the number of spikes is proportional to the amplitude of the left eye saccade (**c**) but not the right eye saccade (**d**). Both cells discharged if the preferred eye moved toward the recording site and were silent if it moved in other directions. Circles indicate data from conjugate saccade trials; inverted triangles indicate data from convergence saccade trials; and upright triangles indicate data from divergence trials. The thin line (**a**) is the regression for the conjugate data (grey circles); the thick lines (**a**, **c**) are the regressions for all of the data.

Finally, if these cells encode conjugate eye movement, then the regression coefficients for right and left eye should be identical. This was not true for any monocular EBN ($P > 0.05$). To illustrate the distribution of eye preference in the EBN population, we calculated an index of burst ocular selectivity, which is similar to eye position ocular selectivity. Monocular EBNs have ocular selectivities near ± 1 (Fig. 3b, lower panel, black bars). In contrast to motor neurons, however, monocular EBNs were evenly distributed in eye preference (37 ipsilateral and 37 contralateral).

We also recorded 65 NPH neurons that encoded eye position. Like EBNs but in contrast to motor neurons, 78% of the NPH cells were monocular, most of them related to the ipsilateral eye. The most parsimonious interpretation of the predominance of monocular units in the PPRF and NPH is that premotor neurons encode commands for monocular eye movements.

These data indicate a new and quite unexpected organization of a motor system: motor neurons that innervate a single eye may exhibit activity related to movements of either eye, whereas premotor cells, predicted according to Hering's law to encode signals related to movements of both eyes, actually exhibit activity related to movements of one eye. Thus, the basic organization of the oculomotor system is probably monocular, and may be related to an evolutionary inheritance of lateral eyes that move independently. The existence of binocular motor neurons, however, indicates that convergence of premotor monocular signals may be crucial for binocular coordination. The distribution of ocular selectivities in the motor neuron pool may also provide a target for adaptation of ocular alignment, with important consequences for the aetiology and management of abnormal eye alignment (strabismus)¹¹. □

Methods

We trained three macaque monkeys to binocularly track a visible target (a laser spot) whose position was controlled by mirror galvanometers. Movements of the left and right eyes were dissociated by aligning the target with one eye and then moving it towards (convergence trials) or away from (divergence trials) the monkey. We assessed the relationship between firing rate and eye movement quantitatively, using multiple regression. Multiple regression analyses were done using a four- (motor neurons and NPH cells) or two- (EBNs) variable linear model: for motor neurons, firing rate was related to the position and speed of the left and right eyes; for EBNs, the number of spikes during the burst was related to the amplitude of left and right eye saccades. For the purpose of brevity, only the data relating to eye position or saccade amplitude are reported here. Eye movements were recorded binocularly using an electromagnetic search coil technique. Animals were prepared for chronic microelectrode recording. We obtained abducens nerve recordings from VIth nerve rootlets ventral to the abducens nucleus and we verified these recordings by microstimulation. EBNs were recorded ventral and rostral to the abducens nucleus.

Received 7 August 1997; accepted 27 March 1998.

- Helmholtz, J. A. *Treatise on Physiological Optics* (Dover, New York, 1910).
- Hering, E. *The Theory of Binocular Vision* (Plenum, New York, 1977).
- Henn, V., Hepp, K. & Büttner-Ennever, J. A. The primate oculomotor system II. Premotor system. *Hum. Neurobiol.* **1**, 87–95 (1982).
- Henn, V. & Cohen, B. Quantitative analysis of activity in eye muscle motoneurons during saccadic eye movements and positions of fixation. *J. Neurophysiol.* **36**, 115–126 (1973).
- Henn, V. & Cohen, B. Coding of information about rapid eye movements in the pontine reticular formation of alert monkeys. *Brain Res.* **108**, 307–325 (1976).
- Luschei, E. S. & Fuchs, A. F. Activity of brainstem neurons during eye movements of alert monkeys. *J. Neurophysiol.* **35**, 445–461 (1972).
- Fuchs, A. F. & Luschei, E. S. Firing patterns of abducens neurons of alert monkeys in relationship to horizontal eye movement. *J. Neurophysiol.* **33**, 382–392 (1970).
- Mays, L. E. & Porter, J. D. Neural control of vergence eye movements: activity of abducens and oculomotor neurons. *J. Neurophysiol.* **52**, 743–761 (1984).
- Fuchs, A. F., Kaneko, C. R. S. & Scudder, C. A. Brainstem control of saccadic eye movements. *Annu. Rev. Neurosci.* **8**, 307–337 (1985).
- Keller, E. L. Participation of medial pontine reticular formation in eye movement generation in monkey. *J. Neurophysiol.* **37**, 316–332 (1974).
- King, W. M. & Zhou, W. An ocular selectivity model of eye movement control: implications for strabismus. *Soc. Neurosci. Abstr.* **23**, 1559 (1997).

Acknowledgements. We thank J. Cai for writing the data acquisition software; C. Drake for care of our animals; B.-F. Tang and J. Alison for technical support; and M. Goldberg and P. May for comments on a preliminary version of this manuscript. This research was supported by NIH and ONR grants to W.M.K.

Correspondence and requests for materials should be addressed to W.Z. (e-mail: wuz@vor.umsd.edu).

Silent glutamatergic synapses and nociception in mammalian spinal cord

Ping Li & Min Zhuo

Department of Anesthesiology, Campus Box 8054, Department of Anatomy and Neurobiology, Washington University in St Louis, St Louis, Missouri 63110, USA

Neurons in the superficial dorsal horn of the spinal cord are important for conveying sensory information from the periphery to the central nervous system^{1,2}. Some synapses between primary afferent fibres and spinal dorsal horn neurons may be inefficient or silent³. Ineffective sensory transmission could result from a small postsynaptic current that fails to depolarize the cell to threshold for an action potential or from a cell with a normal postsynaptic current but an increased threshold for action potentials. Here we show that some cells in the superficial dorsal horn of the lumbar spinal cord have silent synapses: they do not respond unless the holding potential is moved from -70 mV to $+40$ mV. Serotonin (5-hydroxytryptamine, 5-HT), an important neurotransmitter of the raphe–spinal projecting pathway, transforms silent glutamatergic synapses into functional ones. Therefore, transformation of silent glutamatergic synapses may serve as a cellular mechanism for central plasticity in the spinal cord.

Silent glutamatergic synapses have been reported in various regions of the central nervous system (CNS)^{4–9}. Glutamate is a major fast neurotransmitter in the superficial dorsal horn of the spinal cord^{10–12}. We recorded currents from sensory neurons in the superficial dorsal horn of spinal cord slices using whole-cell patch-clamp techniques to test for the existence of silent glutamatergic

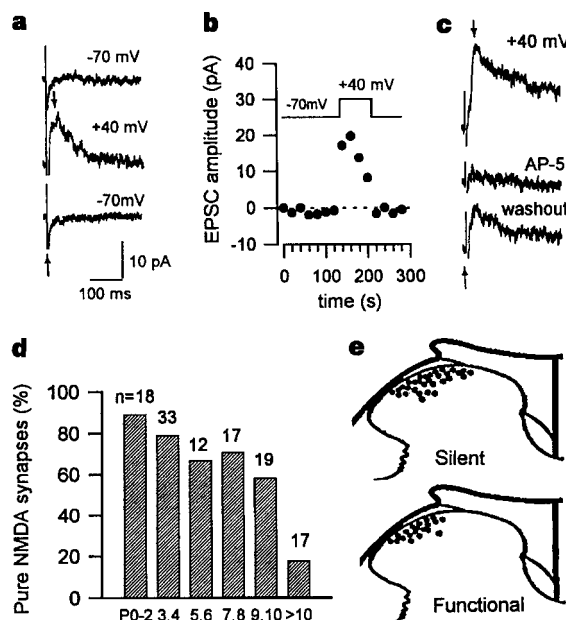


Figure 1 Silent glutamatergic synapses in the lumbar spinal cord. **a**, Examples of responses (the average of three continuous traces) at -70 or $+40$ mV holding potential. **b**, Time course of the experiment shown in **a**. **c**, Responses at $+40$ mV were inhibited by 50μ M AP-5. Upward arrow indicates the time of stimulation. Downward arrow indicates the peak currents measured. **d**, The percentage of silent synapses in the superficial dorsal horn of the lumbar spinal cord at P2–17. The numbers of cells tested are indicated above the bars. **e**, Distribution of labelled spinal neurons²⁸.

synapses. Fast monosynaptic, excitatory postsynaptic currents (EPSCs) were induced when cells from postnatal rats (days postnatal (P) 2–17) were held at -70 mV. To detect silent glutamatergic synapses, we decreased the intensity of stimulation so that no fast EPSC was detected at -70 mV. We then changed the holding potential to $+40$ mV to detect possible *N*-methyl-D-aspartate (NMDA)-receptor-mediated EPSCs. In 68 of 116 experiments, we found synaptic responses at $+40$ mV. After changing the holding potential back to -70 mV, we detected no fast EPSC (Fig. 1a, b). Synaptic responses at $+40$ mV holding potential were abolished by the selective NMDA-receptor antagonist AP-5 ($50 \mu\text{M}$) (Fig. 1c). The pharmacological effect of AP-5 and the voltage dependence indicate that synaptic responses measured at $+40$ mV are NMDA-receptor-mediated EPSCs. The intensity of stimulation used in silent experiments was low and it is likely that only low-threshold afferent fibres are activated. We find that these cells also receive inputs from high-threshold afferent fibres (P.L. and M.Z., unpublished data). However, our techniques could not determine whether silent synapses might be made by high-threshold afferent fibres.

Silent synapses were more abundant in P2–10 than P11–17 rat dorsal horns (Fig. 1d) as in the hippocampus⁷. Silent synapses disappeared from the somatosensory cortex⁶ by P8–9, but were still present in the lumbar spinal cord during P8–10. We detected no anatomical separation of neurons that contained silent or functional synapses (Fig. 1e). Silent synapses can be regulated by activity-dependent synaptic potentiation, known as long-term potentiation^{4–7} (LTP). It is unclear whether the transformation of silent synapses into functional ones strictly requires activation of NMDA receptors—other signalling pathways, such as those involving G-protein-coupled receptors, have not been investigated.

Sensory transmission in the spinal cord receives descending modulation from supraspinal structures including the rostroventral medulla (RVM)^{13–15}. 5-HT-containing neurons in the RVM send

descending projecting terminals to the spinal cord, including the superficial dorsal horn. Activation of these descending pathways can facilitate or inhibit spinal nociceptive transmission^{13–17}. 5-HT at high doses produces inhibition, whereas a selective 5-HT₂-receptor agonist facilitates fast EPSCs in the lumbar spinal cord¹². We carried out a series of experiments to determine whether 5-HT could activate silent synapses in the lumbar spinal cord. We found that 5-HT at low doses could facilitate EPSCs (Fig. 2). To exclude the involvement of NMDA receptors, AP-5 ($50 \mu\text{M}$) was included in the bath solution. 5-HT at high doses inhibited fast EPSCs (Fig. 2). However, a facilitatory effect was induced by 5-HT at low doses and persisted during washout of 5-HT (Fig. 2). Prolonged application (up to 30 min) of 5-HT did not cause further enhancement ($n = 4$; results not shown).

To avoid the activation of fibres from spinal glutamate-containing interneurons¹⁸, we stimulated attached dorsal root nerves. Fast EPSCs induced by dorsal root stimulation were significantly inhibited by a selective 5-HT_{1A}-receptor agonist, 8-hydroxy-2-dipropyl-aminotetralin (8-OH-DPAT; $10 \mu\text{M}$; $n = 5$; Fig. 2). The inhibitory effect of 8-OH-DPAT was blocked by pretreatment with a selective 5-HT_{1A}-receptor antagonist 1-(2-methoxyphenyl)-4-[4-(2-[phthalimido]butyl)-piperazine] HBr (NAN-190) (Fig. 2). By contrast, 1-(2,5-dimethoxy-4-idophenyl)-2-aminopropane (DOI) ($10 \mu\text{M}$), a selective 5-HT₂-receptor agonist, significantly facilitated fast EPSCs ($n = 6$; Fig. 2). The facilitatory effect produced by DOI was prevented by a 5-HT-receptor antagonist methysergide ($1 \mu\text{M}$; $n = 3$). Application of methysergide after DOI failed to reverse the facilitatory effect ($n = 3$).

In the somatosensory cortex, silent synapses disappear by P8–9, the time at which the susceptibility of these synapses to LTP is lost⁶. In the lumbar spinal cord where silent synapses were found, 5-HT at low doses enhanced EPSCs. This effect of 5-HT might be caused by recruiting silent glutamatergic synapses. Application of 5-HT

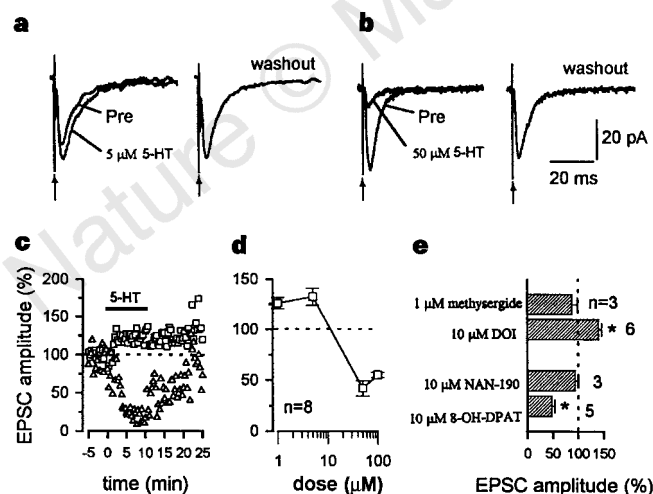


Figure 2 Biphasic modulation induced by 5-HT. **a, b**, Examples of 5-HT experiments at two doses. Upward arrows indicate the time of stimulation. **c**, The effect of 5-HT on amplitudes of EPSCs in experiments shown in **a** (squares) and **b** (triangles). **d**, Summary data of 5-HT at four different doses ($n = 8$ for each dose). **e**, Different effects of 5-HT_{1A}- and 5-HT₂-receptor agonists (8-OH-DPAT and DOI, respectively). The effects were blocked by their receptor antagonists NAN-190 (5-HT_{1A}) and methysergide (5-HT), respectively. * $P < 0.05$.

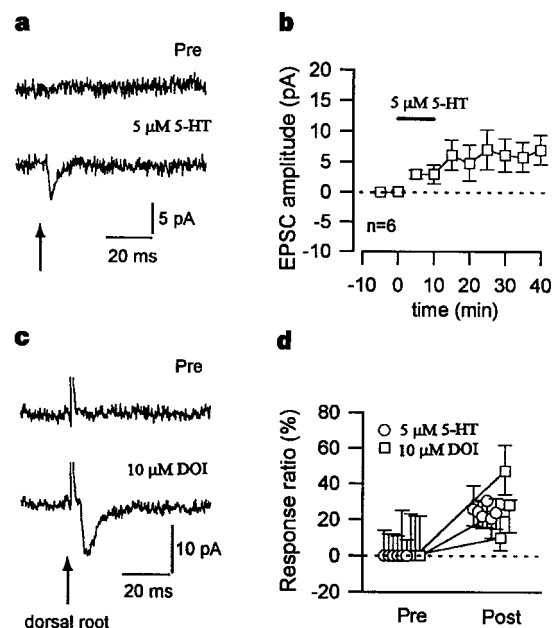


Figure 3 Transformation of silent synapses into functional synapses by 5-HT or DOI. **a**, The average of 15 continuous responses (collected at 0.05 Hz) before and 15 min after the end of 5-HT application. **b**, Summary data of 5-HT. **c**, The average of 15 continuous responses (collected at 0.05 Hz) before and 15 min after the end of DOI application using dorsal root nerve stimulation. **d**, Summary of response ratios before and after 5-HT or DOI application. Symbols for a given cell are connected; symbols and errors show response ratio (the number of stimuli with responses/total number of stimuli $\times 100$) and the calculated 95% confidence intervals²⁹.

(5 μ M) caused typical fast EPSCs to appear (Fig. 3). Saline (see Methods) did not induce any synaptic response ($n = 3$). In experiments using dorsal root stimulation, DOI (10 μ M; $n = 3$) also induced fast EPSCs, as did 5-HT (Fig. 3).

Experiments using minimal stimulation techniques were also carried out¹⁹. 5-HT (5 μ M, 10 min) produced a long-lasting enhancement of fast EPSCs (Fig. 4a). The failure (that is, lack of response) rate at -70 mV holding potential before 5-HT application ($n = 6$, $74 \pm 7\%$) was significantly greater than that during ($59 \pm 12\%$) or after ($44 \pm 9\%$; $P < 0.05$) 5-HT application. It was difficult to distinguish failures from responses at $+40$ mV.

5-HT could affect spinal sensory transmission by acting on presynaptic and/or postsynaptic receptors^{12,20}. To distinguish between these mechanisms, we used G-protein inhibitors applied postsynaptically through the recording pipette. Guanosine 5'-O-(2-thiophosphate) (GDP- β -S) (2 mM) prevented the enhancement produced by 5-HT (5 μ M, $n = 5$, Fig. 4). Slow synaptic responses mediated by substance P or neurokinin were also blocked under these conditions (results not shown), suggesting that GDP- β -S effectively inhibits G proteins. A hydrolysis-resistant GTP analogue, guanosine 5'-O-3-thiotriphosphate (GTP- γ -S; 0.5 mM), also blocked the facilitatory effect of 5-HT ($n = 3$). As a test for the requirement of a postsynaptic Ca^{2+} signalling pathway, we used a pipette solution containing 11 mM BAPTA (1,2-bis(*o*-aminophenoxy)ethane-*N,N,N',N'*-tetra-acetic acid). The facilitatory effect of 5-HT was abolished ($n = 6$; Fig. 4), indicating that an increase in postsynaptic Ca^{2+} is required.

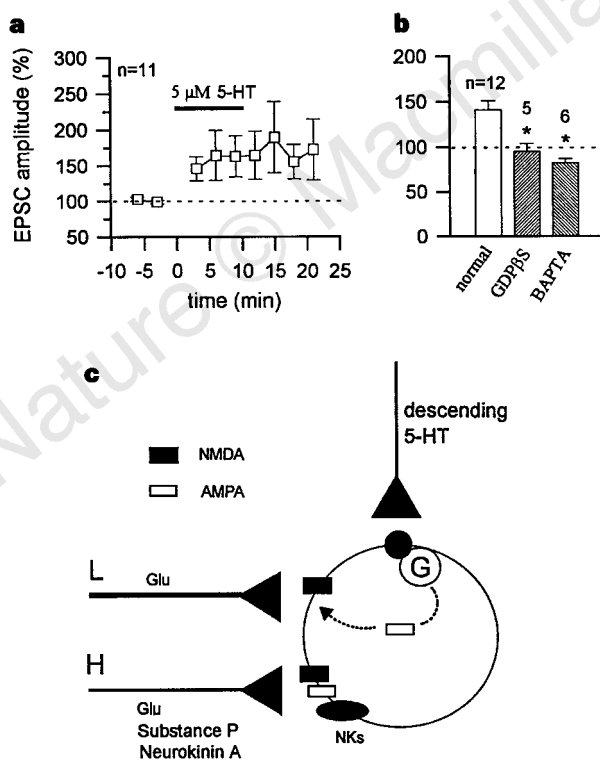


Figure 4 Dependence on postsynaptic G proteins and intracellular Ca^{2+} . **a**, Enhancement by 5-HT using minimal stimulation techniques. Results from five experiments were collected using stimulation at greater intensities. 5-HT gave consistent enhancement and data are pooled. **b**, Postsynaptic application of GDP- β -S (2 mM) or BAPTA (11 mM) abolished the enhancement (10 min after 5-HT). 5-HT enhanced EPSCs when recordings were made with normal pipette solutions. $*P < 0.05$. **c**, Model of silent synapses. A cell receives innervation from low-threshold (L) and high-threshold (H) afferent fibres. Some L inputs are ineffective owing to lack of postsynaptic AMPA receptors. Glu, glutamate; NKS, neurokinin receptors.

Our studies provide direct electrophysiological evidence that ineffective sensory transmission³ may be due to silent glutamatergic synapses in the superficial dorsal horn neurons. Recruitment of silent synapses by activation of postsynaptic G-protein-coupled 5-HT receptors could contribute to enhancement of synaptic transmission induced by 5-HT. As most serotonergic fibres in the spinal dorsal horn originate from the RVM^{13,14}, recruitment of silent synapses by 5-HT could enhance neuronal activity in synapses that receive descending serotonergic signals and may thus promote formation of the descending modulatory neuronal circuit in the spinal cord by an activity-dependent mechanism^{21,22}.

Silent synapses may be involved in persistent pain. Two types of cells in the spinal cord respond to noxious stimuli: wide dynamic range (WDR) cells (which respond to both noxious and non-noxious stimuli) and nociceptive-specific cells (which respond to noxious stimuli only). Recruitment of silent synapses on a WDR cell could enhance its response to sensory stimuli, particularly non-noxious stimuli. For nociceptive-specific cells, activation of silent glutamatergic synapses might even change its phenotype and cause it to respond to non-noxious stimuli (Fig. 4). These synaptic mechanisms could contribute to plastic changes in nociception after tissue or nerve injury^{23–25}. An increase in α -amino-3-hydroxy-5-methyl-4-isoxalepropionate (AMPA) receptors in the spinal dorsal horn after nerve injury has been reported²⁶. Therefore, understanding the cellular and molecular mechanisms by which transformation of silent receptors occurs in the spinal cord will be helpful for designing drugs to stop persistent pain. □

Methods

Slice preparation. Spinal slices were made from P2–17 rats (Sprague-Dawley, Harlan) using a Vibratome. Slices (150–200 μ m) were kept in a filtering funnel containing oxygenated saline (in mM: NaCl, 113; KCl, 3; NaHCO_3 , 25; NaH_2PO_4 , 1; CaCl_2 , 2; MgCl_2 , 1; D-glucose, 25) at room temperature for at least 2 h.

Electrophysiology. Whole-cell recordings were made using 5–10 M Ω electrodes without fire-polishing. Recording electrodes contained (in mM): 110 caesium methanesulphonate, 5 MgCl₂, 1 EGTA, 40 Na-HEPES, 2 Mg-ATP, 0.1 Na₃-GTP (pH 7.2). The osmolarity was adjusted to 295–300 mOsm. Biocytin (5%) was added to label the cell. Membrane potential was clamped at -70 mV (liquid junction potential not corrected). Series resistance was 15–40 M Ω and monitored throughout the experiments. Postsynaptic EPSCs were evoked at 0.02–0.05 Hz with a bipolar tungsten electrode placed near the central end of the dorsal root. Monosynaptic EPSCs were identified using two criteria¹²: the response latency did not change with increasingly intense electrical stimulation, and EPSCs followed high-frequency stimulation (50 Hz) with reduced amplitude and without changing response latency. Dorsal root nerves were directly stimulated in some experiments. Minimal stimulation techniques were also used^{5,19}. The stimulation artefact was subtracted using an average of visually identified failures or responses after 10 μ M CNQX (6-cyano-7-nitroquinoxaline-2,3-dione). A response or failure was visually identified according to two criteria^{5,19}: a failure trace was not visually different from background noise; and a response trace had a constant response latency and shape like typical fast EPSCs. Bicuculline methiodide (10 μ M) and strychnine hydrochloride (1 μ M) were added to the perfusion solution. Statistical comparisons were made with the use of one-way analyses of variance (ANOVAs: Dunnett test for post-hoc comparison) or Student's *t*-test. $P < 0.05$ was considered significant.

Biocytin staining. Neurons were stained as described²⁷. Briefly, slices were fixed in 1% paraformaldehyde (1–7 d at 4 $^{\circ}\text{C}$). Biocytin staining was achieved using avidin/horseradish peroxidase (Vectastain ABC). Slices were mounted and dehydrated on gelatin-coated glass slides.

Received 3 February; accepted 3 April 1998.

1. Kumazawa, T. & Perl, E. R. Excitation of marginal and substantia gelatinosa neurons in the primate spinal cord: indications of their place in dorsal horn functional organization. *J. Comp. Neurol.* **177**, 417–434 (1978).
2. Light, A. R., Trevino, D. L. & Perl, E. R. Morphological features of functionally defined neurons in the marginal zone and substantia gelatinosa of the spinal dorsal horn. *J. Comp. Neurol.* **186**, 151–171 (1979).

3. Wall, P. D. The presence of ineffective synapses and the circumstances which unmask them. *Phil. Trans. R. Soc. Lond. B* **278**, 361–372 (1977).
4. Liao, D., Hessler, N. A. & Malinow, R. Activation of postsynaptically silent synapses during pairing-induced LTP in CA1 region of hippocampal slice. *Nature* **375**, 400–404 (1995).
5. Isaac, J. T. R., Nicoll, R. A. & Malenka, R. C. Evidence for silent synapses: implications for the expression of LTP. *Neuron* **15**, 427–434 (1995).
6. Isaac, J. T. R., Crair, M. C., Nicoll, R. A. & Malenka, R. C. Silent synapses during development of thalamocortical inputs. *Neuron* **18**, 269–280 (1997).
7. Durand, G. M., Kovalchuk, Y. & Konnerth, A. Long-term potentiation and functional synapse induction in developing hippocampus. *Nature* **381**, 71–75 (1996).
8. Wu, G.-Y., Malinow, R. & Cline, H. T. Maturation of a central glutamatergic synapse. *Science* **274**, 972–976 (1996).
9. Malenka, R. C. & Nicoll, R. A. Silent synapses speak up. *Neuron* **19**, 473–476 (1997).
10. Yoshimura, M. & Jessell, T. M. Amino acid-mediated EPSPs at primary afferent synapses with substantia gelatinosa neurones in the rat spinal cord. *J. Physiol. Lond.* **430**, 315–335 (1990).
11. Randic, M., Jiang, M. C. & Cerne, R. Long-term potentiation and long-term depression of primary afferent neurotransmission in the rat spinal cord. *J. Neurosci.* **13**, 5228–5241 (1993).
12. Hori, Y., Endo, K. & Takahashi, T. Long-lasting synaptic facilitation induced by serotonin in superficial dorsal horn neurones of the rat spinal cord. *J. Physiol. Lond.* **492**, 867–876 (1996).
13. Gebhart, G. F. & Randich, A. In *Brainstem Mechanisms of Behaviour* (eds Klemm, W. R. & Vertes, R. P.) 315–352 (Wiley, New York, 1990).
14. Fields, H. L., Heinricher, M. M. & Mason, P. Neurotransmitters in nociceptive modulatory circuits. *Annu. Rev. Neurosci.* **14**, 219–245 (1991).
15. Zhuo, M. & Gebhart, G. F. Spinal serotonin receptors mediate descending facilitation of a nociceptive reflex from the nuclei reticularis gigantocellularis and gigantocellularis pars alpha in the rat. *Brain Res.* **550**, 35–48 (1991).
16. Zhuo, M. & Gebhart, G. F. Characterization of descending facilitation and inhibition of spinal nociceptive transmission from the nuclei reticularis gigantocellularis and gigantocellularis pars alpha in the rat. *J. Neurophysiol.* **67**, 1599–1614 (1992).
17. Zhuo, M. & Gebhart, G. F. Biphasic modulation of spinal nociceptive transmission from the medullary raphe nuclei in the rat. *J. Neurophysiol.* **78**, 746–758 (1997).
18. Todd, A. J., Spike, R. C., Price, R. F. & Neilson, M. Immunocytochemical evidence that neurotension is present in glutamatergic neurons in the superficial dorsal horn of the rat. *J. Neurosci.* **14**, 774–784 (1994).
19. Stevens, C. F. & Wang, Y. Changes in reliability of synaptic function as a mechanism for plasticity. *Nature* **371**, 704–707 (1994).
20. Lopez-Garcia, J. A. & King, A. E. Pre- and post-synaptic actions of 5-hydroxytryptamine in the rat lumbar dorsal horn *in vitro*: implications for somatosensory transmission. *Eur. J. Neurosci.* **8**, 2188–2197 (1996).
21. Singer, W. Development and plasticity of cortical processing architectures. *Science* **270**, 758–764 (1995).
22. Katz, L. C. & Shatz, C. J. Synaptic activity and the construction of cortical circuits. *Science* **274**, 1133–1138 (1996).
23. Menetrey, D. & Besson, J. M. Electrophysiological characteristic of dorsal horn cells in rats with cutaneous inflammation resulting from chronic arthritis. *Pain* **13**, 343–364 (1982).
24. Hylden, J. L. K., Nahin, R. L., Traub, R. J. & Dubner, R. Expansion of receptive fields of spinal lamina I projection neurons in rats with unilateral adjuvant-induced inflammation: the contribution of dorsal horn mechanisms. *Pain* **37**, 229–243 (1989).
25. Dubner, R. & Ruda, M. A. Activity-dependent neuronal plasticity following tissue injury and inflammation. *Trends Neurosci.* **15**, 96–103 (1992).
26. Harris, J. A., Corsi, M., Quartaroli, M., Arban, R. & Bentivoglio, M. Upregulation of spinal glutamate receptors in chronic pain. *Neuroscience* **14**, 7–12 (1996).
27. Maccaferri, G. & McBain, C. J. Long-term potentiation in distinct subtypes of hippocampal nonpyramidal neurons. *J. Neurosci.* **16**, 5334–5343 (1996).
28. Paxinos, G. & Watson, C. *The Rat Brain in Stereotaxic Coordinates* (Compact 3rd edn CD-Rom; Academic, San Diego, 1997).
29. Beyer, W. H. (ed.) *Handbook of Tables for Probability and Statistics* 2nd edn. (Chemical Rubber Company, Cleveland, 1968).

Acknowledgements. We thank R. W. Tsien and J. H. Steinbach for reviewing the manuscript, N. Gautam for discussion, P. Mason for communication of unpublished data, and G. Maccaferri, J. Li, E. T. Kavalali and X. D. Yang for technical advice. This work was supported in part by a grant from the NIDA.

Correspondence and requests for materials should be addressed to M.Z. (e-mail: zhuom@morpheus.wustl.edu).

Synaptic laminin prevents glial entry into the synaptic cleft

Bruce L. Patton, Arlene Y. Chiu* & Joshua R. Sanes

Department of Anatomy and Neurobiology, Washington University School of Medicine, 660 South Euclid Avenue, Box 8108, St Louis, Missouri 63110, USA
* Division of Neuroscience, Beckman Research Institute of the City of Hope, Duarte, California 91010, USA

Presynaptic and postsynaptic membranes directly oppose each other at chemical synapses, minimizing the delay in transmitting information across the synaptic cleft. Extrasynaptic neuronal surfaces, in contrast, are almost entirely covered by processes from glial cells¹. The exclusion of glial cells from the synaptic cleft, and the long-term stability of synapses, presumably result in large part from the tight adhesion between presynaptic and postsynaptic elements^{2,3}. Here we show that there is another requirement for

synaptic maintenance: glial cells of the skeletal neuromuscular synapse, Schwann cells, are actively inhibited from entering the synaptic cleft between the motor nerve terminal and the muscle fibre. One inhibitory component is laminin 11, a heterotrimeric glycoprotein that is concentrated in the synaptic cleft⁴. Regulation of an inhibitory interaction between glial cells and synaptic cleft components may contribute to synaptic rearrangements, and loss of this inhibition may underlie the loss of synapses that results from injury to the postsynaptic cell^{5–12}.

A basal lamina ensheaths each muscle fibre and occupies the synaptic cleft between nerve and muscle at the skeletal neuromuscular junction (Fig. 1a). Synaptic portions of the basal lamina contain components that organize differentiation of both presynaptic and

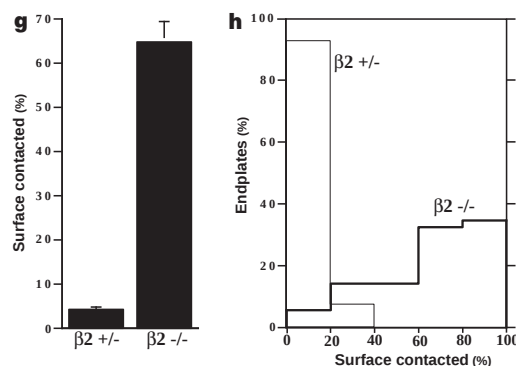
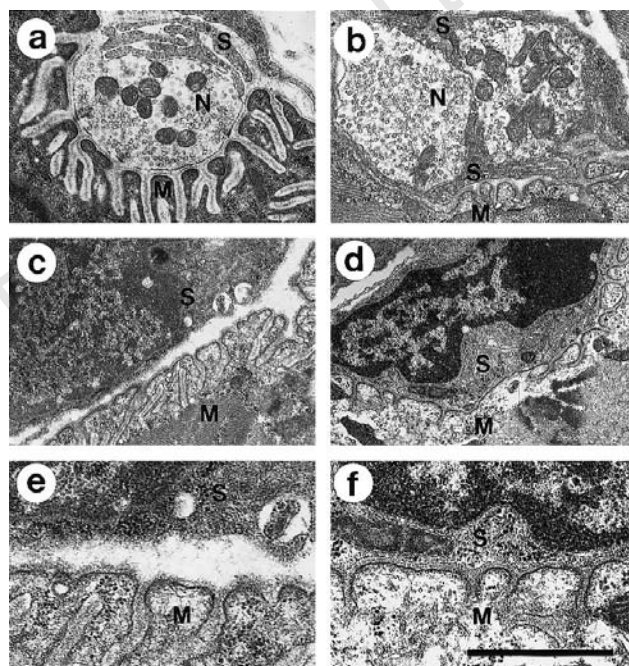


Figure 1 Relationship of Schwann cells to the synaptic cleft at innervated and denervated synaptic sites in control (a, c, e) and laminin $\beta 2^{-/-}$ mutant (b, d, f) muscles. a, b, Schwann cell processes (S) cap nerve terminals (N) on muscle fibres (M) in controls, but extend through the synaptic cleft in mutants. c–f, Following denervation, Schwann cells replace nerve terminals at synaptic sites. Direct contacts between the Schwann cells and the synaptic basal lamina were sparse in controls, but extensive in mutants. Higher-magnification images from c and d are shown in e and f, respectively. Bar (shown in f) represents 1.5 μ m for a, b, 2.0 μ m for c, d, and 0.75 μ m for e, f panels. g, h, Quantification of the direct apposition between Schwann cells and denervated synaptic basal laminae. g, Bars show the fraction of total endplate surface contacted (mean $\pm$ s.d.) from four sternomastoid muscles, examined at days 23–25 postpartum, four days after denervation. h, Data from 123 controls and 93 mutant endplates.

postsynaptic elements¹³. One such component is laminin 11 (refs 4, 14). Laminins are heterotrimers of α , β , and γ subunits; 11 distinct trimers, assembled from five α , three β , and two γ subunits, have been identified to date¹⁵. In adult muscle fibres, all basal laminae are rich in $\gamma 1$ subunits, but the laminins of extrasynaptic basal laminae contain the $\alpha 2$ and $\beta 1$ subunits whereas the $\alpha 5$ and $\beta 2$ subunits are restricted to synaptic basal laminae^{4,16}. Thus, laminin 11 ($\alpha 5\beta 2\gamma 1$) is a prominent synaptic laminin, and laminin 2 ($\alpha 2\beta 1\gamma 1$) is the major extrasynaptic laminin. Mice with a targeted mutation of the laminin $\beta 2$ gene lack both $\alpha 5$ and $\beta 2$ subunits in the synaptic cleft, have malformed neuromuscular junctions, and die in the first postnatal month^{4,17}.

Schwann cell processes, which cap normal neuromuscular junctions, intrude into the synaptic cleft at mutant junctions (Fig. 1a, b), impairing synaptic transmission¹⁷. As laminin $\beta 2$ is adhesive towards motor neurons^{18,19}, we originally thought that nerve terminals failed to adhere tightly to the synaptic basal lamina in $\beta 2$ -laminin-deficient ($\beta 2^{-/-}$) mutants, secondarily allowing Schwann cell processes to invade. It was also possible, however, that the wild-type synaptic cleft actively inhibited Schwann cell entry. To test this idea, we examined denervated muscles, in which nerve terminals had degenerated. Schwann cells enwrap and engulf nerve terminals by phagocytosis following axotomy, and they then occupy the synaptic cleft^{20,21}. However, electron microscopy (Fig. 1c, e) and re-examination of published micrographs^{20,21} showed that Schwann cell processes form few direct contacts with synaptic basal laminae. In contrast, Schwann cells made extensive contacts with synaptic

basal laminae in laminin $\beta 2^{-/-}$ mutants (Fig. 1d, f). Direct apposition of Schwann cells to basal lamina was ~ 15 -fold more prevalent at mutant synaptic sites than at wild-type sites and there were large areas of direct contact at most mutant junctions (Fig. 1g, h). Thus, even in the absence of the nerve, interactions of Schwann cells with the synaptic cleft are dependent on laminin $\beta 2$.

The abnormally close apposition of Schwann cell processes to mutant synaptic cleft material could result from defects in synaptic basal laminae or in the Schwann cell. To distinguish between these alternatives, we plated wild-type Schwann cells on cryostat sections (cryocultures²²) of wild-type or $\beta 2^{-/-}$ lung. We chose lung because it is rich in basal laminae that contain laminin 11 (ref. 15). Schwann cells extended significantly longer processes on mutant lung than on wild-type lung (Fig. 2). Thus, wild-type Schwann cells exhibit distinct behaviours on $\beta 2$ -rich and $\beta 2$ -deficient matrices.

The simplest interpretation of these data is that a $\beta 2$ -containing laminin inhibits process extension by Schwann cells. However, it was also possible that the inhibitory component was bound by laminin. In addition, laminin $\beta 2$ localizes with three laminin α chains in synaptic basal laminae, presumably forming three different trimers (laminins 4 ($\alpha 2\beta 2\gamma 1$), 9 ($\alpha 4\beta 2\gamma 1$) and 11 ($\alpha 5\beta 2\gamma 1$); ref. 4), any or all of which might affect Schwann cells^{23,24}. We therefore examined the behaviour of Schwann cells plated on purified laminins 1 ($\alpha 1\beta 1\gamma 1$), 2, 4, and 11 (laminin 9 was not available). By 12 h after plating, cells had extended processes on laminins 1, 2 and 4, but remained rounded on laminin 11 (Fig. 3a–d). This difference remained marked at 36 h after plating (data not

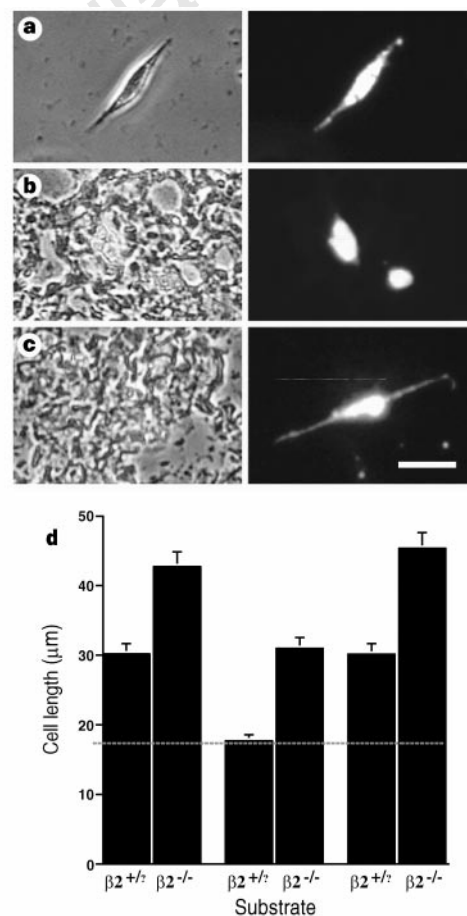


Figure 2 Schwann cells distinguish laminin- $\beta 2$ -deficient from wild-type matrix. **a–c**, DiI-labelled Schwann cells were plated on laminin 1 (**a**) or on cryostat sections of control (**b**) or laminin $\beta 2^{-/-}$ (**c**) lung. The cells spread extensively on laminin $\beta 2^{-/-}$ matrix, as they do on pure laminin 1, but spread poorly on control matrix. Left panels, phase optics; right panels, the same fields viewed with

fluorescence optics. Bar represents $50\mu\text{m}$. **d**, Results from three separate experiments. Each bar represents mean ($\pm$ s.e.m.) of 71–365 (mean = 211) Schwann cells. Dotted line indicates mean lengths of Schwann cells on a non-adhesive substrate, BSA. Values from mutants differ significantly ($P < 0.01$) Student's t -test) from control values in each experiment.

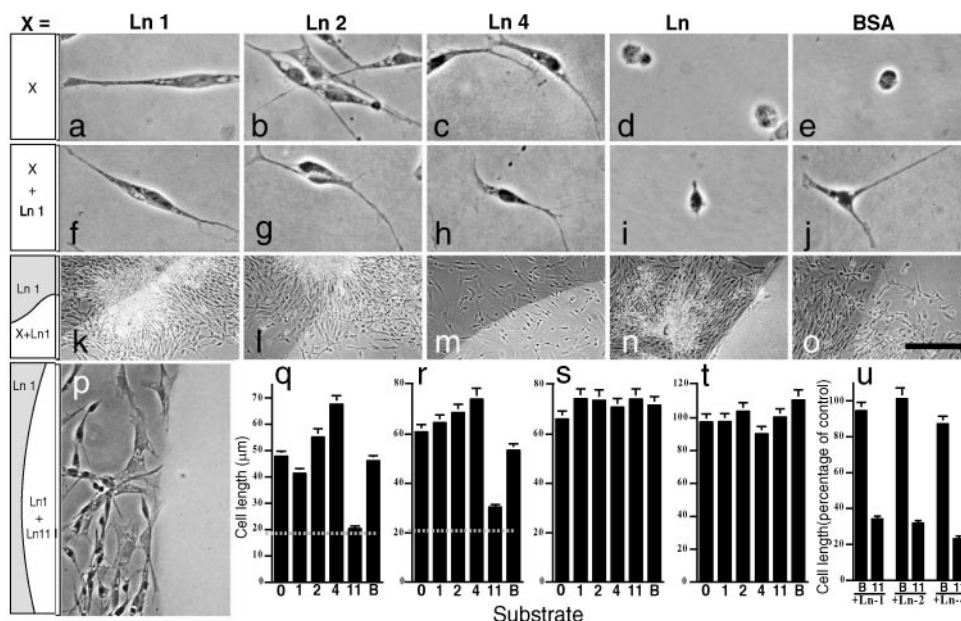


Figure 3 Laminin 11 inhibits process extension by Schwann cells. **a–j**, Schwann cells were plated on substrates coated with test proteins alone (**a–e**) or with mixtures of test proteins plus laminin 1 (**f–j**). **k–p**, Aggregates of Schwann cells were positioned on laminin 1, near borders of test substrate, where they adhered and spread. Spot borders are visible because the test protein was supplemented with a fluorophore, and the field photographed with combined fluorescence and bright-field illumination. The test proteins were laminin 1 (Ln1, **a**, **f**, **k**), laminin 2 (**b**, **g**, **l**), laminin 4 (**c**, **h**, **m**), laminin 11 (**d**, **i**, **n**, **p**) and BSA (**e**, **j**, **o**) (100 µg ml⁻¹). Only laminin 11 inhibited migration promoted by laminin 1. Similar results were obtained for all test proteins at 25 µg ml⁻¹ (not

shown). Bar shown in **o** represents 30 µm for **a–j**, 200 µm for **k–o** and 75 µm for **p**. **q–t**, Process length of Schwann cells (**q**), RN22 Schwannoma cells (**r**), C6 glioma cells (**s**) or QT6 fibroblasts (**t**) plated on mixtures of laminin 1 plus test substrates as in **f–j**. Test substrates were laminin 1 (1), laminin 2 (2), laminin 4 (4), laminin 11 (11), BSA (B) or vehicle alone (0). The dotted line in **q**, **r** indicates the mean cell length on BSA. Bars show mean ± s.e.m. Laminin 11 inhibits process extension by Schwann and Schwannoma cells but not by C6 cells or fibroblasts. **u**, Process length of Schwann cells plated on mixtures of laminins 1, 2 or 4, plus either laminin 11 (11) or BSA (B). In each case, 100% represents the length on unmixed substrate.

shown). Thus, Schwann cells distinguish synaptic laminin 11 from other laminins.

We used two assays to determine whether laminin 11 actively inhibited or simply failed to promote the extension of Schwann cell processes. First, we measured processes on mixtures of laminins. Schwann cells extended processes on mixtures of laminins 1 plus 2 or 1 plus 4, but not on laminins 1 plus 11 (Fig. 3f–j, q), indicating that laminin 11 is inhibitory. Laminin 11 also inhibited process extension promoted by laminins 2 and 4 (Fig. 3u) which, as mentioned above, are present in extrasynaptic and synaptic basal laminae, respectively^{4,16}. In contrast, process extension was normal on mixtures containing laminins 1, 2 or 4 and the non-adhesive substrate bovine serum albumin (BSA) (Fig. 3j, q, u). Second, we measured migration of Schwann cells on patterned substrates^{4,19}. Schwann cells migrated on laminin 1 and readily crossed borders onto spots of laminins 1 plus 2, 1 plus 4, or 1 plus BSA (Fig. 3k–m, o). However, migration stopped abruptly at borders where the Schwann cells encountered laminin 11 (Fig. 3n, p), as occurs at neuromuscular junctions *in vivo*. Inhibition of extension of Schwann cell processes by laminin 11 provides a plausible basis for the exclusion of Schwann cell processes from the synaptic cleft.

We showed previously that laminin 11 stops outgrowth of neurites from motor neurons⁴. To determine whether laminin 11 is generally inhibitory to process outgrowth, we assayed RN22 schwannoma, C6 glioma, and QT6 fibroblast cells. Laminin 11 inhibited spreading of RN22, but not of C6 or QT6 cells (Fig. 3r–t). Thus, laminin 11 exerts inhibitory effects on both primary and immortalized Schwann cells, but not on all cells. We also compared responses of neurons and Schwann cells to laminin β2 fragments. Inhibition of neurite outgrowth is due at least in part to a tripeptide of amino-acid sequence/LRE, located near the carboxy terminus of

laminin β2: outgrowth is inhibited by an LRE-containing β2 fragment but not by a fragment in which LRE was mutated to QRE^{18,19}. Spreading and migration of Schwann cells, in contrast, was inhibited equally well by the wild-type and mutant fragments (Fig. 4a–c). Furthermore, concatamers of an LRE-containing decapeptide¹⁹ inhibited neurite outgrowth but not Schwann cell spreading or migration (Fig. 4d–f; data not shown). The different specificities of neurons and Schwann cells indicate that different receptors may mediate responses of these two cell types.

We have found that glia are actively inhibited from entering the synaptic cleft; this inhibition promotes the close apposition of presynaptic and postsynaptic cells. Thus, synaptic stability results not only from adhesive forces between synaptic partners but also from the synaptic cleft's ability to repel glial invasion. Glial cells exert several stimulatory effects on the formation and function of synapses^{25–27}; the inhibitory effect that we describe here provides a means of ensuring that trophically advantageous neuron–glia interactions do not become so intimate that they disrupt synaptic communication. Conversely, loss of inhibitory components could contribute to the ensheathing of motor nerve terminals by Schwann cells that follows damage to muscle fibres^{5,6}. Similarly, glial processes separate nerve terminals from synaptic sites following axotomy of both central and peripheral neurons^{7–12}, a phenomenon called 'synaptic stripping'. The mechanisms that underlie synaptic stripping are unknown; our results raise the possibility that altered interactions between glial cells and synaptic cleft components may be involved. Finally, it is possible that alterations in synaptic size and distribution, which occur during maturation (synapse elimination) and in adulthood (learning), could result not only from altered interactions between presynaptic and postsynaptic elements²⁸ but also from modulation of the inhibitory effect of cleft components on glial cells. □

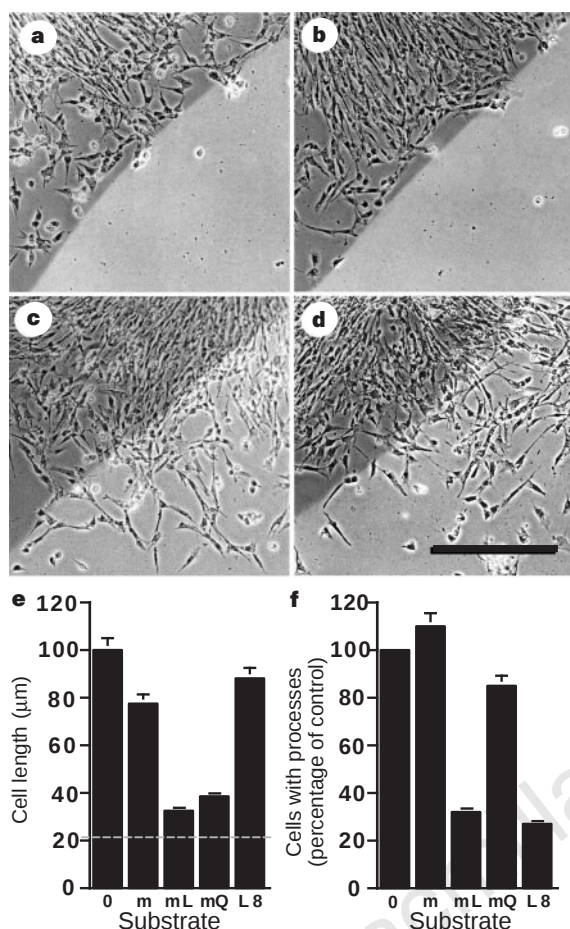


Figure 4 Differential inhibition of neurites and Schwann cells by fragments of laminin $\beta 2$. **a–d**, Schwann cells plated on laminin 1 did not migrate onto mixtures of laminin 1 plus a recombinant laminin $\beta 2$ fragment (**a**), or a mutant fragment in which the LRE sequence had been changed to QRE (**b**), but did migrate onto mixtures of laminin 1 plus preparations from control bacteria (**c**) or an 8-fold concatamer of an LRE-containing decapeptide (**d**)¹⁹. Bar represents 80 μm . **e**, Spreading of Schwann cells on laminin 1 was inhibited by the wild-type (mL) and mutant (mQ) fragments of laminin $\beta 2$ but not by control preparations (m) or by the LRE-containing decapeptide (L8). In all cases 20 $\mu\text{g ml}^{-1}$ laminin 1 was mixed with 100–200 $\mu\text{g ml}^{-1}$ test protein. **f**, Outgrowth of ciliary neurites, in contrast, was inhibited by the wild-type fragment and the decapeptide but not by the QRE mutant¹⁹. In **e, f**, 0 represents values on laminin 1 alone. Dotted line in **e** shows the average cell length on BSA.

Methods

Animals. We generated laminin $\beta 2^{-/-}$ mice as described¹⁷ and maintained them on a high-fat diet to extend lifespan to ~4 weeks⁴. Sternomastoid muscles were denervated by resection of the nerve proximal to the muscle and prepared for electron microscopy¹⁷. Cross-sections were scanned systematically in the electron microscope, and all synaptic sites encountered (defined by the presence of junctional folds and postsynaptic membrane thickening) were photographed at a magnification of $\times 14,000$. The distance between Schwann cell and muscle fibre plasma membranes was measured from micrographs. The closest appositions encountered were ~50 nm, reflecting the presence of the basal lamina. Regions of apposition of ≤ 70 nm were categorized as direct.

Cultures. We dissociated Schwann cells from desheathed sciatic nerves of chick embryos at embryonic day 14 (SPAFAS, Roanoke, IL) by digestion with 0.05% trypsin for 30 min at 37°C in PBS. More than 90% of the viable cells in the resulting suspensions were Schwann cells, as assessed by staining for two cell-type-specific antigens, GFAP glial fibrillary acidic protein and S-100. For the

experiments shown in Fig. 2, cells were plated on uncoated culture dishes, washed, cultured for 6–10 h in 50 $\mu\text{g ml}^{-1}$ DiI (Molecular Probes), washed and replated on 14- μm -thick cryostat sections, which were prepared as described²². For experiments shown in Figs 3k–p and 4, Schwann cells were aggregated in hanging drops as described for neurons in ref. 15; the preformed aggregates were then transferred with a pipette to patterned substrates, and positioned near borders. Cells were fixed after 24–36 h in culture. Cell lengths were measured from digital images acquired with a charge-coupled device video camera (Photometrics; Tucson, AZ) and IPLab software (Signal Analysis, Vienna, VA). RN22 schwannoma, C6 glioma and QT6 fibroblastic cells were obtained from the American Type Culture Collection. For experiments shown in Figs 3 and 4, substrates were coated with laminins or laminin fragments as described^{17,21}. Laminins 2 and 4 were a gift from Y.-S. Cheng and P. Yurchenco²⁹. Laminin 11 was prepared as in ref. 4, and laminin $\beta 2$ fragments were prepared as in ref. 19.

Received 5 February; accepted 26 March 1998.

- Peters, A., Palay, S. L. & Webster, H. D. *The Fine Structure of the Nervous System* (Oxford Univ. Press, New York, 1991).
- Vaughn, J. E. Fine structure of synaptogenesis in the vertebrate central nervous system. *Synapse* **3**, 255–285 (1989).
- Colman, D. R. Neurites, synapses, and cadherins reconciled. *Mol. Cell. Neurosci.* **10**, 1–6 (1997).
- Patton, B. L., Miner, J. H., Chiu, A. Y. & Sanes, J. R. Localization, regulation and function of laminins in the neuromuscular system of developing, adult and mutant mice. *J. Cell Biol.* **139**, 1507–1521 (1997).
- Duchen, L. W., Excell, B. J., Patel, R. & Smith, B. Changes in motor end-plates resulting from muscle fibre necrosis and regeneration. A light and electron microscopic study of the effects of the depolarizing fraction (cardiotoxin) of *Dendroaspis jamesoni* venom. *J. Neurol. Sci.* **21**, 391–417 (1974).
- Jirmanová, I. Ultrastructure of motor end-plates during pharmacologically-induced degeneration and subsequent regeneration of skeletal muscle. *J. Neurocytol.* **4**, 141–155 (1975).
- Matthews, M. R. & Nelson, V. H. Detachment of structurally intact nerve endings from chromatolytic neurones of rat superior cervical ganglion during the depression of synaptic transmission induced by post-ganglionic axotomy. *J. Physiol.* **245**, 91–135 (1975).
- Purves, D. Functional and structural changes in mammalian sympathetic neurones following interruption of their axons. *J. Physiol.* **252**, 429–463 (1975).
- Blinzinger, K. & Kreutzberg, G. Displacement of synaptic terminals from regenerating motoneurons by microglial cells. *Z. Zellforschung* **85**, 145–157 (1968).
- Svensson, M. & Aldskogius, H. Synaptic density of axotomized hypoglossal motoneurons following pharmacological blockade of the microglial cell proliferation. *Exp. Neurol.* **120**, 123–131 (1993).
- Titmus, M. J. & Faber, D. S. Axotomy-induced alterations in the electrophysiological characteristics of neurons. *Progr. Neurobiol.* **35**, 1–51 (1990).
- Kreutzberg, G. W. In *The Axon: Structure, Function and Pathophysiology* (eds Waxman, J. D. K. & Stys, P. K.) 355–374 (Oxford Univ. Press, New York, 1995).
- Hall, Z. W. & Sanes, J. R. Synaptic structure and development: the neuromuscular junction. *Cell* **72**, 99–121 (1993).
- Hunter, D. D., Shah, V., Merlie, J. P. & Sanes, J. R. A laminin-like adhesive protein concentrated in the synaptic cleft of the neuromuscular junction. *Nature* **338**, 229–234 (1989).
- Miner, J. H. *et al.* The laminin alpha chains: expression, developmental transitions, and chromosomal locations of $\alpha 1$ – $\alpha 5$, identification of heterotrimeric laminins 8–11, and cloning of a novel $\alpha 3$ isoform. *J. Cell Biol.* **137**, 685–701 (1997).
- Noakes, P. G., Gautam, M., Mudd, J., Sanes, J. R. & Merlie, J. P. Aberrant differentiation of neuromuscular junctions in mice lacking s-laminin/laminin $\beta 2$. *Nature* **374**, 258–262 (1995).
- Hunter, D. D. *et al.* Primary sequence of a motor neuron-selective adhesive site in the synaptic basal lamina protein S-laminin. *Cell* **59**, 905–913 (1989).
- Porter, B. E., Weiss, J. & Sanes, J. R. A motoneuron-selective stop signal in the synaptic protein s-laminin. *Neuron* **14**, 549–559 (1995).
- Miledi, R. & Slater, C. R. Electrophysiology and electron-microscopy of rat neuromuscular junctions after nerve degeneration. *Proc. R. Soc. Lond. B* **169**, 289–306 (1968).
- Winlow, W. & Usherwood, P. N. R. Ultrastructural studies of normal and degenerating mouse neuromuscular junctions. *J. Neurocytol.* **4**, 377–394 (1975).
- Covault, J., Cunningham, J. M. & Sanes, J. R. Neurite outgrowth on cryostat sections of innervated and denervated skeletal muscle. *J. Cell Biol.* **105**, 2479–2488 (1987).
- Anton, E. S., Sandrock, A. W. Jr & Matthew, W. D. Merosin promotes neurite growth and Schwann cell migration *in vitro* and nerve regeneration *in vivo*: evidence using an antibody to merosin, ARM-1. *Dev. Biol.* **164**, 133–146 (1994).
- Milner, R. *et al.* Division of labor of Schwann cell integrins during migration on peripheral nerve extracellular matrix ligands. *Dev. Biol.* **185**, 215–228 (1997).
- Son, Y.-J., Trachtenberg, J. T. & Thompson, W. J. Schwann cells induce and guide sprouting and reinnervation of neuromuscular junctions. *Trends Neurosci.* **19**, 280–285 (1996).
- Pfrieffer, F. W. & Barres, B. A. New views on synapse–glia interactions. *Curr. Opin. Neurobiol.* **6**, 615–621 (1996).
- Pfrieffer, F. W. & Barres, B. A. Synaptic efficacy enhanced by glial cells *in vitro*. *Science* **277**, 1684–1688 (1997).
- Bargmann, C. Making memories stick? *Nature* **391**, 435–436 (1998).
- Cheng, Y.-S., Champlaud, M.-F., Burgeson, R. E., Marenkovich, M. P. & Yurchenco, P. D. Self-assembly of laminin isoforms. *J. Biol. Chem.* **272**, 31525–31532 (1997).

Acknowledgements. We thank J. Cunningham and S. Weng for assistance, Y.-S. Cheng and P. Yurchenco for laminins, J. Lichtman for comments, and J. Liu for participation in preliminary experiments. This work was supported by grants from the NIH and the NSF.

Correspondence and requests for materials should be addressed to J.R.S. (e-mail: sanesj@thalamus.wustl.edu).

Association of missense and 5'-splice-site mutations in *tau* with the inherited dementia FTDP-17

Mike Hutton^{*1}, Corinne L. Lendon^{*2}, Patrizia Rizzu^{*3,4}, Matt Baker¹, Susanne Froelich^{3,5}, Henry Houlden¹, Stuart Pickering-Brown⁶, Sumi Chakraverty², Adrian Isaacs¹, Andrew Grover¹, Jennifer Hackett¹, Jennifer Adamson¹, Sarah Lincoln¹, Dennis Dickson¹, Peter Davies⁷, Ronald C. Petersen⁸, Martijn Stevens⁴, Esther de Graaf³, Erwin Wauters³, Jeltje van Baren³, Marcel Hillebrand³, Marijke Joosse³, Jennifer M. Kwon⁹, Petra Nowotny², Lien Kuei Che², Joanne Norton⁹, John C. Morris⁵, Lee A. Reed¹⁰, John Trojanowski¹⁰, Hans Basun⁵, Lars Lannfelt⁵, Michael Neystat¹¹, Stanley Fahn¹¹, Francis Dark¹², Tony Tannenberg¹³, Peter R. Dodd¹⁴, Nick Hayward¹⁵, John B. J. Kwok¹⁶, Peter R. Schofield¹⁶, Athena Andreadis¹⁷, Julie Snowden¹⁸, David Craufurd¹⁹, David Neary¹⁸, Frank Owen⁶, Ben A. Oostra³, John Hardy¹, Alison Goate², John van Swieten⁴, David Mann²⁰, Timothy Lynch¹¹ & Peter Heutink³

^{*} These authors contributed equally to this work

Thirteen families have been described with an autosomal dominantly inherited dementia named frontotemporal dementia and parkinsonism linked to chromosome 17 (FTDP-17)¹⁻⁹, historically termed Pick's disease¹⁰. Most FTDP-17 cases show neuronal and/or glial inclusions that stain positively with antibodies raised against the microtubule-associated protein Tau, although the Tau pathology varies considerably in both its quantity (or severity) and characteristics^{1-8,12}. Previous studies have mapped the FTDP-17 locus to a 2-centimorgan region on chromosome 17q21.11; the *tau* gene also lies within this region. We have now sequenced *tau* in FTDP-17 families and identified three missense mutations (G272V, P301L and R406W) and three mutations in the 5' splice site of exon 10. The splice-site mutations all destabilize a potential stem-loop structure which is probably involved in regulating the alternative splicing of exon 10 (ref. 13). This causes more frequent usage of the 5' splice site and an increased proportion of *tau* transcripts that include exon 10. The increase in exon 10⁺ messenger RNA will increase the proportion of Tau containing four microtubule-binding repeats, which is consistent with the neuropathology described in several families with FTDP-17 (refs 12, 14).

FTDP-17 is a condition characterized clinically by behavioural, cognitive and motor disturbance¹. At autopsy all patients with FTDP-17 have pronounced frontotemporal atrophy with loss of neuronal cells, grey and white matter gliosis and superficial cortical spongiform changes. Variable Tau inclusions are observed in the brains of most FTDP-17 patients. Tau is also the major component of the paired helical filaments¹⁵ that make up the characteristic tangles seen in the brains of patients with Alzheimer's disease and with other neurodegenerative disorders. The Tau isoforms that predominate in human brain are encoded by eleven exons¹¹. A

total of six different major *tau* mRNA transcripts are generated as a result of alternative splicing and encode proteins of 352–441 amino acids¹⁶. Alternative splicing of exon 10 generates Tau protein with 3 or 4 microtubule-binding motifs that are imperfect repeats of 31 or 32 residues¹⁵.

Previous studies^{1,3-5} found no evidence of mutations in the *tau* gene that were associated with FTDP-17, but these were done on individual families and were mainly restricted to the coding region of the gene. We have now extended the analysis of the *tau* gene to additional families with FTDP-17 and to other regions of the gene. The 11 *tau* coding exons and flanking intronic regions were initially sequenced in 40 individuals from families with frontotemporal dementia from Scandinavia (9 families)¹⁷, the Netherlands (3 families), the USA (5 families), Australia (1 family; Fig. 1a) and from Greater Manchester in the UK (22 families)¹⁷. Eight of these families had previously displayed evidence for linkage to chromosome 17 (refs 1–3, 7).

We detected two missense mutations (G272V and P301L, numbered from the longest Tau isoform) that occur in two of the microtubule-binding repeat domains of Tau^{15,18}. The P301L mutation (Fig. 1b) in exon 10 was found in two families (Table 1): a large Dutch kindred hereditary frontal temporal dementia I (HFTDI)⁷, and a small kindred from the United States (FTD003). This substitution occurs in a highly conserved region of the Tau sequence, where a proline residue is found in all mammalian species from which *tau* has been cloned so far. The P301L mutation will only affect the 4-repeat Tau isoforms because exon 10 is spliced out of mRNA that encodes the 3-repeat isoforms¹¹. Analysis of Tau aggregates in affected brains from the FTD003 family (P301L) reveals that these consist mainly of 4-repeat isoforms, consistent with a mutation affecting exon 10 (P.D., manuscript in preparation).

The G272V mutation was found in a second large Dutch kindred HFTD2 (Table 1), originally described as having hereditary Pick's disease⁷. This mutation also affects a highly conserved residue within the microtubule-binding domain, encoded by exon 9. Within the imperfect repeat sequence that makes up the four microtubule-binding domains, the G272V and P301L mutations affect positions that are separated by only one residue. Thus, for P301L the invariant PGGG motif in the binding repeat becomes LGGG, and for G272V it becomes PGVG. In contrast to the P301L mutation (exon 10), the G272V mutation (exon 9) will affect all Tau isoforms. The G272V and P301L mutations segregate with disease in each of the relevant families. Both mutations were absent from 192 Dutch controls and the P301L mutation was also absent from 150 US controls. These results indicate that the G272V and the P301L mutations are probably pathogenic.

A third *tau* missense mutation (R406W) was detected in exon 13 in a single family from the United States (FTD004)¹⁹ which alters a highly conserved residue near the carboxy terminus. This mutation segregates with the disease in this family and was absent from 150 US controls. The distribution of Tau-positive inclusions in FTD004

Table 1 Families with segregating mutations in the *tau* gene

Family	Origin (founder)	Affecteds*	Generations	Mean onset age	Mutation
HFTD2†	Netherlands	34(15)	7	47	G272V
HFTD1†	Netherlands	49(14)	5	50	P301L
FTD003	USA	3(2)	2	45–50	P301L
Man19	UK	3(1)	2	65	Ex10 splice + 13
DDPAC†	Ireland	13(7)	3	44	Ex10 splice + 14
Aus1†	Australia (UK)	28(5)	5	53	Ex10 splice + 16
FTD002†	USA	3(1)	2	40	Ex10 splice + 16
Man6	UK	2(1)	1	48	Ex10 splice + 16
Man23†	UK	10(2)	3	51	Ex10 splice + 16
FTD004	USA	10(2)	4	55	R406W

* Confirmed post mortem in brackets.

† Families with prior evidence of genetic linkage to chromosome 17.

¹ Mayo Clinic Jacksonville, Jacksonville, Florida 32224, USA; ² Dept of Psychiatry, Washington University School of Medicine, St Louis, Missouri 63110, USA; ³ Dept of Clinical Genetics, ⁴ Dept of Neurology, Erasmus University Rotterdam, The Netherlands; ⁵ Section of Geriatric Medicine, Institution of Clinical Neuroscience and Family Medicine, Karolinska Institute, S-14186 Huddinge, Sweden; ⁶ Divn of Neuroscience, School of Biological Sciences, University of Manchester, Manchester M13 9PT, UK; ⁷ Dept of Pathology, Albert Einstein College of Medicine, New York, New York 10461, USA; ⁸ Mayo Clinic, Rochester, Minnesota 56007, USA; ⁹ Dept of Neurology, Washington University School of Medicine, St Louis, Missouri 63110, USA; ¹⁰ Dept of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104-4283, USA; ¹¹ Dept of Neurology, Columbia University, New York, New York 10032, USA; ¹² Dept of Psychiatry, Princess Alexandra Hospital, Wooloongaba, 4102, Australia; ¹³ Dept of Anatomical Pathology, Mater Misericordiae Hospital, Brisbane, 4101, Australia; ¹⁴ Dept of Biochemistry, University of Queensland, Brisbane, 4072, Australia; ¹⁵ Human Genetics Laboratory, Queensland Institute of Medical Research, Herston, 4029, Australia; ¹⁶ Garvan Institute of Medical Research, Sydney, NSW 2010, Australia; ¹⁷ Dept of Biomedical Sciences, E.K. Shriver Center, Waltham, Massachusetts 02154, USA; ¹⁸ Dept of Neurology, Manchester Royal Infirmary, Manchester M13 9WL, UK; ¹⁹ Dept of Clinical Genetics, St Mary's Hospital, Manchester M13 0GH, UK; ²⁰ Dept of Pathological Sciences, University of Manchester, Manchester M13 9PT, UK.

meets the NINDS neuropathological criteria for progressive supranuclear palsy (PSP)¹⁹. Electron microscopy, however, revealed that the Tau filaments in this family are Alzheimer's-like paired helical filaments¹⁹ and not the straight Tau filaments normally observed in PSP¹². The neuropathological phenotype in this family is also similar to that of the *lytico* and *bodig* diseases of Guam²⁰. It may be important that the R406W mutation lies near key residues that are phosphorylated in Tau from paired helical filaments (Ser 396, Ser 404)²¹.

In addition to the three missense mutations (G272V, P301L and R406W), we also found three heterozygous mutations in a cluster of 4 nucleotides 13–16 base pairs (bp) 3' of the exon-10 5' splice site (Fig. 3). Six families (Table 1) had mutations at these 3 sites, including 4 families who had previously shown linkage to chromosome 17 (DDPAC², Aus1 (ref. 3), FTD002 and Man23). In each of the 6 families, the relevant mutation was found to segregate with the disease (Fig. 1a): none of these intronic variants was observed in 150 US and 23 British caucasian controls. These results indicated that this cluster of mutations (Fig. 2a, b) could be pathogenic, but it was not obvious how such intronic mutations could influence exon-10 splicing and thus the function of Tau.

To test whether these 5'-splice-site mutations could be affecting alternative splicing of exon 10 and thus altering the proportion of *tau* mRNA containing this exon, we used the polymerase chain reaction with reverse transcription (RT-PCR) to estimate the ratio of *tau* exon10⁺ RNA to exon10⁻ RNA. Seven FTDP-17 brains were analysed: four were from families with 5'-splice-site mutations (DDPAC (2 brains)², ManF23 and AusI (refs 3, 9)) and three were from families with the P301L mutation (FTD003 (2 brains) and HFTD1 (ref. 7)). RT-PCR was done between exons 9 and 11 and, in a separate reaction, between exons 9 and 13. Both amplifications generated two products, one corresponding to *tau* transcripts containing exon 10 and one to *tau* transcripts lacking exon 10. In both PCRs the four FTDP-17 brains carrying 5'-splice-site mutations gave a two- to sixfold higher proportion of *tau* exon10⁺ RNA compared with 7 control brains (Fig. 2c). FTDP-17 brains without splice-site mutations (P301L mutants) gave similar ratios to control brains (Fig. 2c), indicating that the increased ratio in brains with the

splice-site mutations is not secondary to the disease process. It needs to be investigated whether the variability between the splice-site mutant brains (two- to sixfold increase over controls) reflects the variable nature of the disease phenotype in these families, particularly the wide range in onset age. Our results indicate that the 5'-splice-site mutations in FTDP-17 families increase the proportion of *tau* mRNA containing exon 10 and encoding the 4-repeat isoforms.

We used exon-trapping assays to test the effect of the 5'-splice-site mutations on alternative splicing of *tau* exon 10 (ref. 22). Wild-type and the three mutant (+13, +14 and +16) versions of exon 10 were analysed. These sequences, including ~40 bp of intronic sequence at either end, were amplified and cloned into the splicing vector pSPL3b, which contains exons from the rabbit β -globin and HIV *tat* genes. A multiple cloning site in an intron between the two *tat* exons allows test DNA to be introduced. An SV40 promoter in pSPL3b drives the generation of artificial mRNAs when the construct is transfected into COS7 cells, trapping any functional exons in the cloned DNA between the two *tat* exons so they can be detected by RT-PCR (Fig. 3).

Exon-trapping analysis²² of wild-type *tau* exon-10 constructs (Fig. 3) gave multiple bands, the most prominent at 177 bp corresponding to vector-only *tat* to *tat* exon splicing (exon10⁻ mRNA). Weaker products at 270 bp corresponded to exon 10⁺ transcripts. In contrast, each of the *tau* exon-10 5' splice mutant constructs gave products mainly of 270 bp, corresponding to exon 10⁺ transcripts. With the mutant constructs, *tat* to *tat* spliced vector-only transcripts (177 bp) were greatly reduced, consistent with increased usage of the exon-10 5' splice site. *tau* exon 10-vector-only transcripts were almost eliminated with the +13 and +14 mutants. These results show that the 5'-splice-site mutants increase splicing of exon 10 *in vitro*.

The mechanism by which the intronic mutations increase exon 10⁺ *tau* mRNA is not fully understood. However, we examined the sequence of the exon-10 5' splice site and found a short potential stem-loop structure (12-base stem, 6-base loop) which spans the splice site (Fig. 2a, b). The three mutations occurred in the stem of this structure and should destabilize it. It has been reported that short stem-loop structures can sequester 5' splice sites and lead to usage of alternative 5' splice sites²³. This suggested that the stem-loop structure could be involved in regulating exon-10 alternative splicing, presumably by competing with the U1 snRNP (or with other factors recognizing this site) for binding to the exon-10 5' splice site²⁴. Blocking U1 snRNP (or another factor) binding would result in failure to define exon 10 (through U1 and U2 snRNP binding) and lead to skipping of this exon²⁴. This would allow the splicing of exon 9 to exon 11 and the generation of *tau* transcripts lacking exon 10. We propose that the ratio of transcripts with and without exon 10 reflects the stability of this stem-loop structure, which in turn determines the ratio of 4-repeat and 3-repeat Tau. Mutations in the stem-loop that are associated with FTDP-17 because of their destabilizing effect on the structure will promote recognition of the 5' splice site²⁴ and the inclusion of exon 10 in *tau* mRNA. This should increase the proportion of *tau* transcripts containing exon 10, just as we found in our RT-PCR analysis of FTDP-17 splice-mutant brains and in the exon-trapping analysis, and hence the ratio of 4-repeat to 3-repeat isoforms. This idea is consistent with the observation that soluble Tau in many FTDP-17 families consists of more of the 4-repeat isoforms than that from control brains¹⁴. The absence of this type of splice-site mutation in other FTDP-17 families (including families with missense mutations) is consistent with the finding that, in at least one FTDP-17 family, the relative abundance of Tau isoforms with 3- and 4-repeats is similar to that of soluble Tau in control brains¹².

Our results have several implications for frontotemporal dementia and neurodegeneration. First, we have identified mutations that cause FTDP-17, one of the major autosomal-dominant

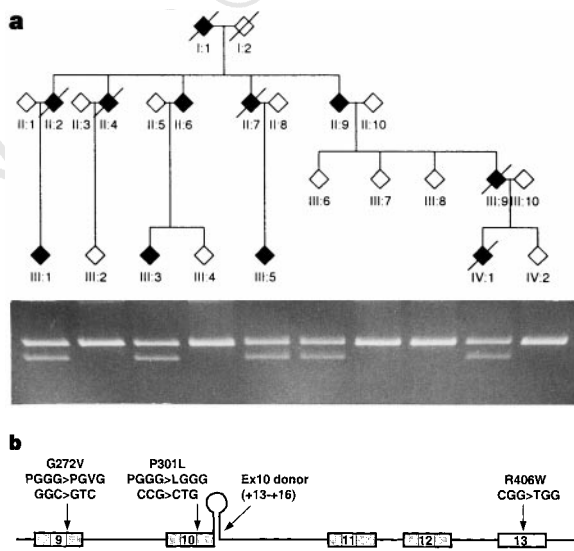


Figure 1 *tau* mutations in FTDP-17. **a**, Segregation analysis of the exon-10 5' splice site +16 mutation in the AusI kindred. The mutation is indicated by a band at 148 bp on the agarose gel whereas the normal allele is represented by a band at 200 bp. Unaffected individual III.6 carries the mutation and disease haplotype. **b**, Diagram of the *tau* gene (exons 9–13) showing locations of missense and 5'-splice-site mutations. Shaded boxes denote exons encoding microtubule-binding domains. The effect of two missense mutations G272V and P301L on the PGGG motif in the microtubule-binding domains is shown.

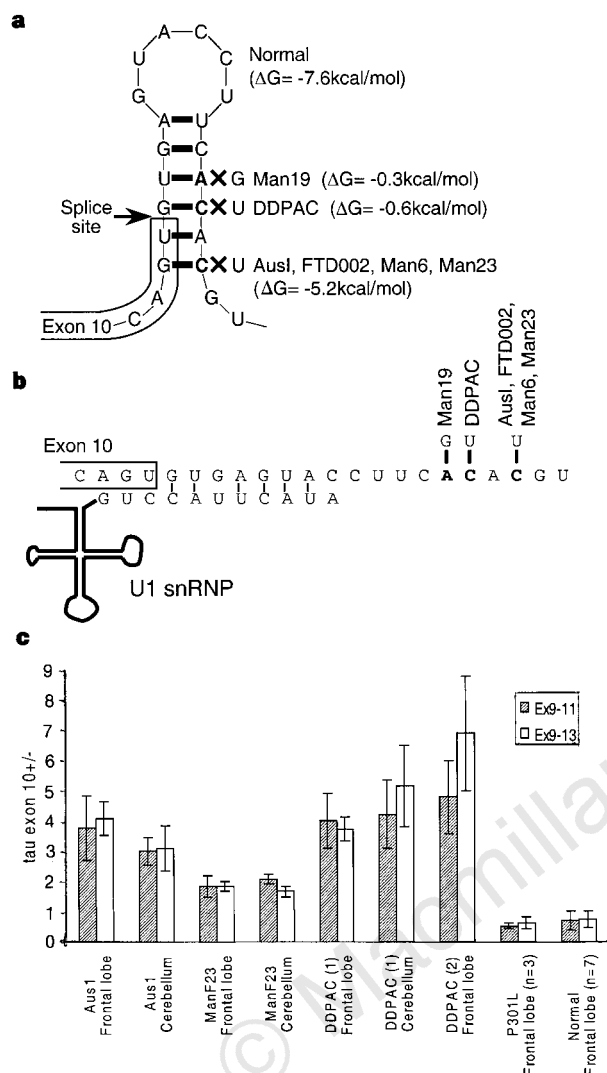


Figure 2 *tau* exon-10 5'-splice-site mutations. The splice site in the **a**, the predicted stem-loop structure, and **b**, in the linear form. Mutations are marked and the predicted free energy of the stem-loop is indicated²⁸. Predicted binding of U1 snRNP to the 5' splice site is shown. **c**, RT-PCR analysis of the molar ratio between *tau* mRNA with and without exon 10. RNA from frontal lobe (4 cases) and cerebellum (3 cases) from FTDP-17 brains with splice mutations (DDPAC (2), ManF23 and Aus1) has ratios >1.6 (left, 14 bars): RNA from frontal lobe of normal brains ($n = 7$) and P301L mutation brains has ratios <0.8 (right, 4 bars). Analysis of total *tau* mRNA revealed no significant differences.

loci associated with neurodegeneration¹. Frontotemporal dementia, of which FTDP-17 is a familial form, accounts for 3–10% of dementia cases¹. Our data also indicate that the recently reported¹² Tau-sequence variant V337M in a family with FTDP-17 (Seattle A) is probably pathogenic. Because no mutations were previously found in other families, it was suggested that this might be a benign polymorphism, although a pathogenic effect in this particular family was not excluded¹². Second, and most important, our identification of pathogenic missense (G272V, P301L and R406W) and 5'-splice-site mutations associated with FTDP-17 shows that Tau dysfunction can lead to neurodegeneration. In addition, the nature of the 5'-splice-site mutations indicates that the relative levels of 4-repeat and 3-repeat isoforms are crucial to the functioning of Tau, consistent with alternative splicing of exon 10 being developmentally regulated¹³. Other tauopathies²⁵, including Pick's disease, PSP, corticobasal degeneration and *lytico* and *bodig* diseases of Guam²⁰, should be tested for evidence of *tau* gene dysfunction—

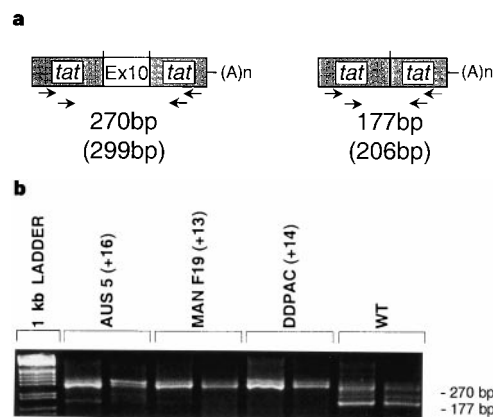


Figure 3 5'-splice-site mutations increase incorporation of *tau* exon 10 into artificial mRNAs. **a**, Diagram of major RT-PCR products generated by exon trapping analysis of *tau* exon 10. PCR products result from splicing of pSPL3b vector-derived *tat* exons with (270 bp) and without (177 bp) *tau* exon 10. In addition, there are minor products of 299 bp and 206 bp that contain an additional 29 bp of vector sequence. The position of nested RT-PCR primers is indicated by arrows. **b**, Results of exon trapping analysis of *tau* exon 10. Wild-type constructs (WT) gave strongest signals from bands (177 bp, 206 bp) corresponding to exon 10⁺ vector-only splicing. Splice-site mutants gave strongest signals corresponding to exon 10⁺ transcripts (270 bp, 299 bp).

an association between an intronic polymorphism in the *tau* gene and PSP has already been reported^{26,27}. Our results also have implications for Alzheimer's disease because Tau paired-helical-filament pathology is a hallmark of this disease. Finally, the nature of the mutations identified in the FTDP-17 families indicate that abnormalities in Tau microtubule binding are probably important in the pathogenesis of this group of tauopathies. It is not clear whether abnormal microtubule binding would lead directly to cell death (through microtubule destabilization), or whether it would merely result in an increase in unbound Tau that could then form pathogenic fibrils which disrupt cellular function. □

Methods

***tau* gene sequencing.** *tau* exons (1–4,5,7,9–13) were amplified from genomic DNA from family members with primers designed to flank intronic sequence. PCRs contained a final concentration of 0.8 μM for each primer and 1 unit of *Taq* 'Gold' polymerase (Perkin Elmer). Amplification was done using a 60–50 °C touchdown protocol over 35 cycles with a final extension of 72 °C for 10 min. PCR products were purified using the Qiagen PCR kit and their concentration estimated on an agarose gel. DNA (100 ng) for each exon was sequenced on both strands using the Rhodamine dye terminator cycle sequencing kit (Perkin Elmer) and relevant PCR primers. Sequencing was performed on an ABI377 automated sequencer. Heterozygote base calls were made using Factura software (Perkin Elmer) and sequence alignment was performed with Sequence Navigator (Perkin Elmer).

Mutation detection. Mutations were detected in families and controls using sequencing or by PCR–RFLP analysis. Missense mutation P301L was detected by *Bst*NI or *Sma*I digestion of exon-10 PCR product, the mutant allele contains a *Bst*NI site, the normal allele a *Sma*I site. The G272V mutation was detected by *Acc*II digestion of exon 9 PCR products, the mutation eliminates the *Acc*II site. The R406W missense mutation was detected by *Nco*I digestion of exon 13 PCR product, the mutant allele contains the *Nco*I site. The exon 10 splice donor +16 mutation was detected by *Nsp*I digestion, again the mutant allele is cleaved by this assay. In contrast, the exon-10 5' splice site +13 and +14 mutations eliminate an *Afl*III site from the amplification product. After digestion, genotyping was done on 3% Metaphor (FMC) agarose gels.

RT-PCR analysis of exon 10 alternative splicing. Total RNA was prepared from a section of frontal lobe from seven normal brains and from the frontal lobes of seven FTDP-17 brains using the Trizol reagent and protocol (Life Technologies). Four of the FTDP-17 brains were from families with 5'-splice-site

mutations (DDPAC (2 brains), ManF23 and Aus1) and 3 were from families with the P301L point mutant (FTD003 (2 brains) and HFTD1). Reverse transcription was performed using the Superscript preamplification kit (Life Technologies) on 1–4 µg of brain RNA with an oligo(dT) primer. PCR was performed between exon 9 (forward, 5'-ATCGCAGCGCTACAGCAG-3') and exon 11 (reverse, 5'-TGGTTTATGATGGATGTTGCC-3') and between exon 9 and exon 13 (reverse 5'-TCTTGGCTTTGGCGTCTC-3'). In each case, the 5' end of the forward amplification primer was labelled with TET (Perkin Elmer) to allow detection by an ABI377 automated sequencer. Preliminary PCRs (not shown) were performed using a range of amplification cycles (18–37) to determine the optimum number of cycles for this analysis. Based on these results, we used 32 cycles in subsequent experiments. After amplification, PCR products were analysed on an ABI377 automated sequencer (Perkin Elmer) where they resolved into two major fragments (327 and 418 bp, exons 9–11; 487 and 578 bp, exon 9–13) corresponding to *tau* transcripts with and without exon 10. The identity of each band was confirmed by sequence analysis. The molar ratio of exon 10⁺ to exon 10⁻ RNA was determined using Genescan software. Three independent PCRs (for both exons 9–11 and 9–13) were used to determine the mean and s.d. of the ratio for each brain. Results (not shown) were essentially identical when exon 10^{+/−} ratios were estimated by densitometric analysis of PCR products on agarose gels using a Kodak DC120 camera kit and ID Image Gel densitometry software.

Exon-trapping analysis of exon-10 splicing. Mutant and wild-type versions of *tau* exon 10 were amplified from the DNA of patients with each of the three different splice mutations (+13, +14 and +16) and from normal individuals. PCR products contained exon 10 and ~40 bp of flanking intron sequence at either end. PCR products were cloned into the splicing vector pSPL3b using *XhoI* and *PstI* sites incorporated into the amplification products. Mutant and wild-type constructs were identified by sequence analysis. For exon trapping, the exon-trapping system of Life Technologies was used. Briefly, COS-7 cells were transfected in duplicate with 1 µg each construct using LipofectACE reagent (Life Technologies). Cells were collected 24 h post-transfection and RNA prepared using the Trizol reagent (Life Technologies). First-strand synthesis and nested PCR were done using reagents supplied with the system and conditions described in manufacturer's instructions, except that *Bst*X1 digestion of primary PCR products was excluded. To verify that the RT-PCR was quantitative, different amounts of primary PCR template (1–5 µl) were used and the total number of amplification cycles was varied (30–35 cycles). PCR products were analysed on 3% Metaphore (FMC) gels. RT-PCR products (Fig. 3) had their identities confirmed by sequencing.

Received 9 April; accepted 1 June 1998.

1. Foster, N. L. *et al.* Frontotemporal Dementia and Parkinsonism Linked to Chromosome 17: A Consensus Statement. *Annu. Neurol.* **41**, 706–715 (1997).
2. Wilhelmsen, K. C., Lynch, T., Pavlou, E. & Nygaard, T. G. Localization of disinhibition-dementia-parkinsonism-amyotrophy complex to 17q21–22. *Am. J. Hum. Genet.* **55**, 1159–1165 (1994).
3. Baker, M. *et al.* Localization of frontotemporal dementia with parkinsonism in an Australian kindred to chromosome 17q21–22. *Annu. Neurol.* **42**, 794–798 (1997).
4. Froelich, S. *et al.* Mapping of a disease locus for familial rapidly progressive frontotemporal dementia to chromosome 17q12–21. *Am. J. Med. Genet.* **74**, 380–385 (1997).
5. Murrell, J. *et al.* Familial multiple system tauopathy with presenile dementia localized to chromosome 17. *Am. J. Hum. Genet.* **61**, 1131–1138 (1997).
6. Wijk, M. *et al.* Localization of the gene for rapidly progressive autosomal dominant parkinsonism and dementia with pallido-ponto-nigral degeneration to chromosome 17q21. *Hum. Mol. Genet.* **5**, 151–154 (1996).
7. Heutink, P. *et al.* Hereditary fronto-temporal dementia is linked to chromosome 17q21–22. A genetic and clinico-pathological study of three Dutch families. *Annu. Neurol.* **41**, 150–159 (1997).
8. Yamaoka, L. H. *et al.* Linkage of frontotemporal dementia to chromosome 17: clinical and neuropathological characterization of phenotype. *Am. J. Hum. Genet.* **59**, 1306–1312 (1996).
9. Dark, F. A family with autosomal dominant, non-Alzheimer's presenile dementia. *Austr. N. Z. J. Psychiat.* **31**, 139–144 (1997).
10. Constantinidis, J., Richard, J. & Tissot, R. Pick's disease: Histological and clinical classification. *Eur. Neurol.* **11**, 208–217 (1974).
11. Andreadis, A., Brown, W. M. & Kosik, K. S. Structure and novel exons of the human *tau* gene. *Biochemistry* **31**, 10626–10633 (1992).
12. Spillantini, M. G., Bird, T. D. & Ghetti, B. Frontotemporal dementia and parkinsonism linked to chromosome 17: A new group of tauopathies. *Brain Path.* **8**, 387–402 (1998).
13. Goedert, M., Spillantini, M. G., Potier, M. C., Ulrich, J. & Crowther, R. A. Cloning and sequencing of the cDNA encoding an isoform of microtubule-associated protein *tau* containing four tandem repeats: differential expression of *tau* protein mRNAs in human brain. *EMBO J.* **8**, 393–399 (1989).
14. Spillantini, M. G. *et al.* Familial multiple system tauopathy with presenile dementia: a disease with abundant neuronal and glial tau filaments. *Proc. Natl Acad. Sci. USA* **94**, 4113–4118 (1997).
15. Goedert, M. *et al.* Assembly of microtubule-associated protein *tau* into Alzheimer-like filaments induced by sulphated glycosaminoglycans. *Nature* **383**, 550–553 (1996).
16. Goedert, M., Spillantini, M. G., Jakes, R., Rutherford, D. & Crowther, R. A. Multiple isoforms of human microtubule-associated protein *tau*: sequence and localization in neurofibrillary tangles of Alzheimer's disease. *Neuron* **3**, 519–526 (1989).

17. The Lund and Manchester groups. Clinical and neuropathological criteria for frontotemporal dementia. *J. Neurol. Neurosurg. Psychiat.* **57**, 416–418 (1994).
18. Butner, K. A. Kirschner, M. W. J. *Tau* protein binds to microtubules through a flexible array of distributed weak sites. *J. Cell. Biol.* **115**, 717–730 (1991).
19. Reed, L. A. *et al.* Autosomal dominant dementia with widespread neurofibrillary tangles. *Annu. Neurol.* **42**, 564–572 (1997).
20. McGeer, P. L., Schwab, C., McGeer, E. G., Haddock, R. L. & Steele, J. C. Familial nature and continuing morbidity of the amyotrophic lateral sclerosis-parkinsonism dementia complex of Guam. *Neurology* **49**, 400–409 (1997).
21. Goedert, M. Tau protein and the neurofibrillary pathology of Alzheimer's disease. *Trends Neurosci.* **16**, 460–465 (1993).
22. Church, D. M. *et al.* Isolation of genes from complex sources of mammalian genomic DNA using exon amplification. *Nature Genet.* **6**, 98–105 (1994).
23. Eperon, I. P., Graham, I. R., Griffiths, A. D. & Eperon, I. C. Effects of RNA secondary structure on alternative splicing of pre-mRNA: is folding limited to a region behind the transcribing RNA polymerase? *Cell* **54**, 393–401 (1988).
24. Kuo, H.-C., Nasim, F.-U. H. & Grabowski, P. J. Control of alternative splicing by the differential binding of U1 small nuclear ribonucleoprotein particle. *Science* **251**, 1045–1050 (1991).
25. Dickson, D. Neurodegenerative diseases with cytoskeletal pathology: a biochemical classification. *Annu. Neurol.* **42**, 541–543 (1997).
26. Conrad, C. *et al.* Genetic evidence of the involvement of τ in progressive supranuclear palsy. *Annu. Neurol.* **41**, 277–281 (1997).
27. Higgins, J. J., Litvan, I., Pho, L. T., Li, W. & Nee, L. E. Progressive supranuclear palsy is in linkage disequilibrium with the τ and not the α -synuclein gene. *Neurology* **50**, 270–273 (1998).
28. Freier, S. M. *et al.* Improved free-energy parameters for predictions of RNA duplex stability. *Proc. Natl Acad. Sci. USA* **83**, 9373–9377 (1986).

Acknowledgements. This work was supported by an NINDS (RO1) grant to M.H., an NIA (MCSDA) grant to T.L. and NIA (P50) grants to A.G. and J.C.M. and to M.H., J.H. and R.C.P. Additional support was provided by the Mayo Foundation (M.H., J.H.), the International Foundation for Alzheimer Research, the Dutch Brain Research Foundation (P.H.) and Judith Mason (P. Dodd). A.G. is the recipient of an NIH career development award (NIA). C.L.L. was a Washington University Alzheimer's Disease Research Center postdoctoral fellow. P.R. is the recipient of a TMG of EU grant. J.M.K. is the recipient of an NSADA award. T.L. is the recipient of Irving Scholar, NARSAD and Parkinson's Disease Foundation awards. We thank the Mayo Clinic Molecular Biology Core Facility for sequencing and acknowledge the support of the Dutch Brain Bank (W.K., R.R.), the Michigan ADRC Brain Bank and the Columbia University Brain Bank. The participation of the families has been crucial.

Correspondence and requests for materials should be addressed to M.H. (e-mail: hutton.michael@mayo.edu) or P.H. (e-mail: heutink@kgen.fgg.eur.nl).

The antigenic structure of the HIV gp120 envelope glycoprotein

Richard Wyatt*, Peter D. Kwong†, Elizabeth Desjardins*, Raymond W. Sweet‡, James Robinson§, Wayne A. Hendrickson† & Joseph G. Sodroski*||

* Department of Cancer Immunology and AIDS, Dana-Farber Cancer Institute, Department of Pathology, Harvard Medical School, and || Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, Massachusetts 02115, USA

† Department of Biochemistry and Molecular Biophysics, Howard Hughes Medical Institute, Columbia University, New York, New York 10032, USA

‡ SmithKline Beecham Pharmaceuticals, 709 Swedeland Road, King of Prussia, Pennsylvania 19406-0939, USA

§ Department of Pediatrics, Tulane University Medical Center, 1430 Tulane Avenue, New Orleans, Louisiana 70112, USA

The human immunodeficiency virus HIV-1 establishes persistent infections in humans which lead to acquired immunodeficiency syndrome (AIDS). The HIV-1 envelope glycoproteins, gp120 and gp41, are assembled into a trimeric complex that mediates virus entry into target cells¹. HIV-1 entry depends on the sequential interaction of the gp120 exterior envelope glycoprotein with the receptors on the cell, CD4 and members of the chemokine receptor family^{2–4}. The gp120 glycoprotein, which can be shed from the envelope complex, elicits both virus-neutralizing and non-neutralizing antibodies during natural infection. Antibodies that lack neutralizing activity are often directed against the gp120 regions that are occluded on the assembled trimer and which are exposed only upon shedding^{5,6}. Neutralizing antibodies, by contrast, must access the functional envelope glycoprotein complex⁷ and typically recognize conserved or variable epitopes near the receptor-binding regions^{8–11}. Here we describe the spatial organization of conserved neutralization epitopes on gp120, using epitope

maps in conjunction with the X-ray crystal structure of a ternary complex that includes a gp120 core, CD4 and a neutralizing antibody¹². A large fraction of the predicted accessible surface of gp120 in the trimer is composed of variable, heavily glycosylated core and loop structures that surround the receptor-binding regions. Understanding the structural basis for the ability of HIV-1 to evade the humoral immune response should assist in the design of a vaccine.

The amino-acid sequence of human and simian immunodeficiency virus gp120 glycoproteins consists of five variable regions (V1–V5) interposed among more conserved regions¹³. Variable regions V1–V4 form exposed loops anchored at their bases by disulphide bonds¹⁴.

Neutralizing antibodies recognize both variable and conserved gp120 structures. The V2 and V3 loops contain epitopes for strain-restricted neutralizing antibodies^{15–17}. More broadly neutralizing antibodies recognize discontinuous, conserved epitopes in three regions of the gp120 glycoprotein (Table 1). In HIV-1-infected humans, the most abundant of these are directed against the CD4-binding site (CD4BS) and block gp120–CD4 interaction^{8,9}. Less common are antibodies against epitopes induced or exposed upon CD4 binding (CD4i)¹⁸. Both CD4i and V3 antibodies disrupt the binding of gp120–CD4 complexes to chemokine receptors^{10,11}. A third gp120 neutralization epitope is defined by a unique monoclonal antibody, 2G12 (ref. 19), which does not efficiently block receptor binding¹¹.

In the accompanying Article¹², we report the X-ray crystal structure of an HIV-1 gp120 core in a ternary complex with two-domain soluble CD4 and the Fab fragment of the CD4i antibody 17b. The gp120 core lacks the V1/V2 and V3 variable loops, as well as amino- and carboxy-terminal sequences, which interact with the gp41 glycoprotein⁶, and is enzymatically deglycosylated^{12,20}. Despite these modifications, the gp120 core binds CD4 and antibodies against CD4BS and CD4i epitopes^{20,21} and thus retains structural integrity. The gp120 core is composed of an inner domain, an outer domain and a third element, the 'bridging sheet'¹² (Fig. 1a). All three structural elements contribute, either directly or indirectly, to CD4 and chemokine-receptor binding¹². We now analyse the organization of the surface of the gp120 core in light of the known antibody responses directed against this exposed viral glycoprotein.

Although generally well conserved compared with the five variable regions, some variability in the surface of the gp120 core is

evident when the sequences of all primate immunodeficiency viruses are analysed. This variability is disproportionately associated with the surface of the outer domain proximal to the V4 and V5 regions and removed from the receptor-binding regions (Fig. 1a–c). The LA, LC, LD and LE surface loops¹² contribute to the variability of this surface. The potential N-linked glycosylation sites present in the gp120 core are concentrated in this variable half of the protein (Fig. 1b, c). The only conserved residues apparent on this relatively variable surface are asparagine 356 and threonine/serine 358, which constitute a complex carbohydrate addition site within the LE loop (Fig. 1b, c). As most carbohydrate moieties may appear as 'self' to the immune system, the extensive glycosylation of the outer domain surface may render it less visible to immune surveillance. This helps to explain why antibodies directed against this gp120 surface have been identified so infrequently.

The receptor-binding regions retained in the gp120 core are well conserved among primate immunodeficiency viruses¹². Also highly conserved is the surface of the inner domain spanned by the $\alpha 1$ helix and located opposite the variable surface described above (Fig. 1d). This surface is likely to interact with gp41 and/or with N-terminal gp120 segments absent from the gp120 core. This inner domain surface and the receptor-binding regions are devoid of glycosylation.

In conjunction with prior mutagenic and antibody competition analyses^{5,6,18–20,22} the gp120 core structure reveals the spatial positioning of the conserved gp120 neutralization epitopes. Although the principal variable loops are either absent (V1/V2 and V3) or poorly resolved (V4) in the gp120 core structure, their approximate positions can be deduced (Fig. 2a). The conserved gp120 neutralization epitopes are discussed in relation to these variable loops and the variable, glycosylated core surface.

CD4i epitopes. The gp120 epitope recognized by the CD4i antibody 17b can be directly visualized in the crystallized ternary complex¹² (Fig. 2b, c). Strands from the gp120 fourth conserved (C4) region and the V1/V2 stem contribute to an antiparallel β -sheet (the 'bridging sheet'; Fig. 1a) that contacts the antibody. Most gp120 residues previously implicated in the formation of the CD4i epitopes¹⁸ (Table 1) are located either within this β -sheet or in nearby structures. With the exception of Thr 202 and Met 434, the gp120 residues in contact with the 17b Fab are highly conserved among HIV-1 isolates (Figs 1c, 2a). The prominent ('male') CDR3 loop of the 17b heavy chain dominates the contacts with gp120, with

Table 1 Conserved epitopes for neutralizing antibodies identified on the gp120 core

Competition group*	Examples of monoclonal antibodies	gp120 amino acids†	Probable mechanism of virus neutralization	Properties	Selected refs
CD4-binding site (CD4BS)	F105 15e 21h 1125h 448D 39.3 IgG1b12 830D	Asn 88 (13), Asp 113 (50), Lys 117 (25), Ser 256 (75), Thr 257 (75), Asn 262 (63), Ala 266 (13), Asp 368 (100), Glu 370 (100), Tyr 384 (13), Lys 421 (50), Trp 427 (25), Asp 457 (13), Pro 470 (25), Asp 474 (13), Met 475 (13), Asp 477 (63), Asp/Leu/ Tyr 482/483/484 (25)	Interference with gp120–CD4 binding	CD4BS antibodies compete with CD4 and with antibodies against CD4i epitopes	8, 9, 22
CD4-induced epitopes (CD4i)	17b 48d	Asn 88, Lys 117, Lys 121, Lys 207, Ser 256, Thr 257, Asn 262, Δ V3, Glu 370, Glu 381, Phe 382, Arg 419, Ile 420, Lys 421, Gln 422, Ile 423, Trp 427, Tyr 435, Pro 438, Met 475	Interference with chemokine-receptor binding	CD4 binding increases exposure of the epitopes as a result of movement of the V2 variable loop	18; C. Rizzuto and J.G.S., submitted
2G12	2G12	Asn 295, Thr 297, Ser 334, Asn 386, Asn 392, Asn 397	Unknown	Antibody binding is dependent upon proper N-linked glycosylation	19

* The gp120 competition groups are defined as in ref. 5.

† The gp120 amino acids are numbered according to the sequence of the HXBc2 (IIIB) gp120 glycoprotein, where residue 1 is the methionine at the amino terminus of the signal peptide. Changes in the amino acids listed resulted in significant reduction in antibody binding to the gp120 glycoprotein (refs 18, 19, 22). Numbers in parentheses indicate the percentage of CD4BS antibodies examined whose binding is decreased by changes in the indicated residue.

additional contacts through the heavy chain CDR2 (ref. 12). Unusually, there are minimal 17b light-chain contacts, leaving a large gap between the gp120 core and most of the 17b light-chain surface. In the complete gp120 glycoprotein, this gap is probably occupied by the V3 loop. This is consistent with the position and orientation of the V3 base on the gp120 core structure¹², the effect of V3 deletions on the binding of CD4i antibodies in the absence of soluble CD4 (ref. 21), the competition of some V3-directed antibodies with CD4i antibodies⁵, and the ability of both antibody

groups to block chemokine-receptor binding^{10,11}. The chemokine-receptor-binding region of gp120 probably consists of elements near or within the 'bridging sheet' and the V3 loop (Fig. 1a), a model that is supported by recent mutagenic analysis (C. Rizzuto *et al.*, submitted).

The V2 loop probably resides on the side of the 17b epitope opposite the V3 loop (Fig. 2a). The V1/V2 loops, which vary from 57 to 86 residues in length¹³, are dispensable for HIV-1 replication^{21,23}, but decrease the sensitivity of viruses to neutraliza-

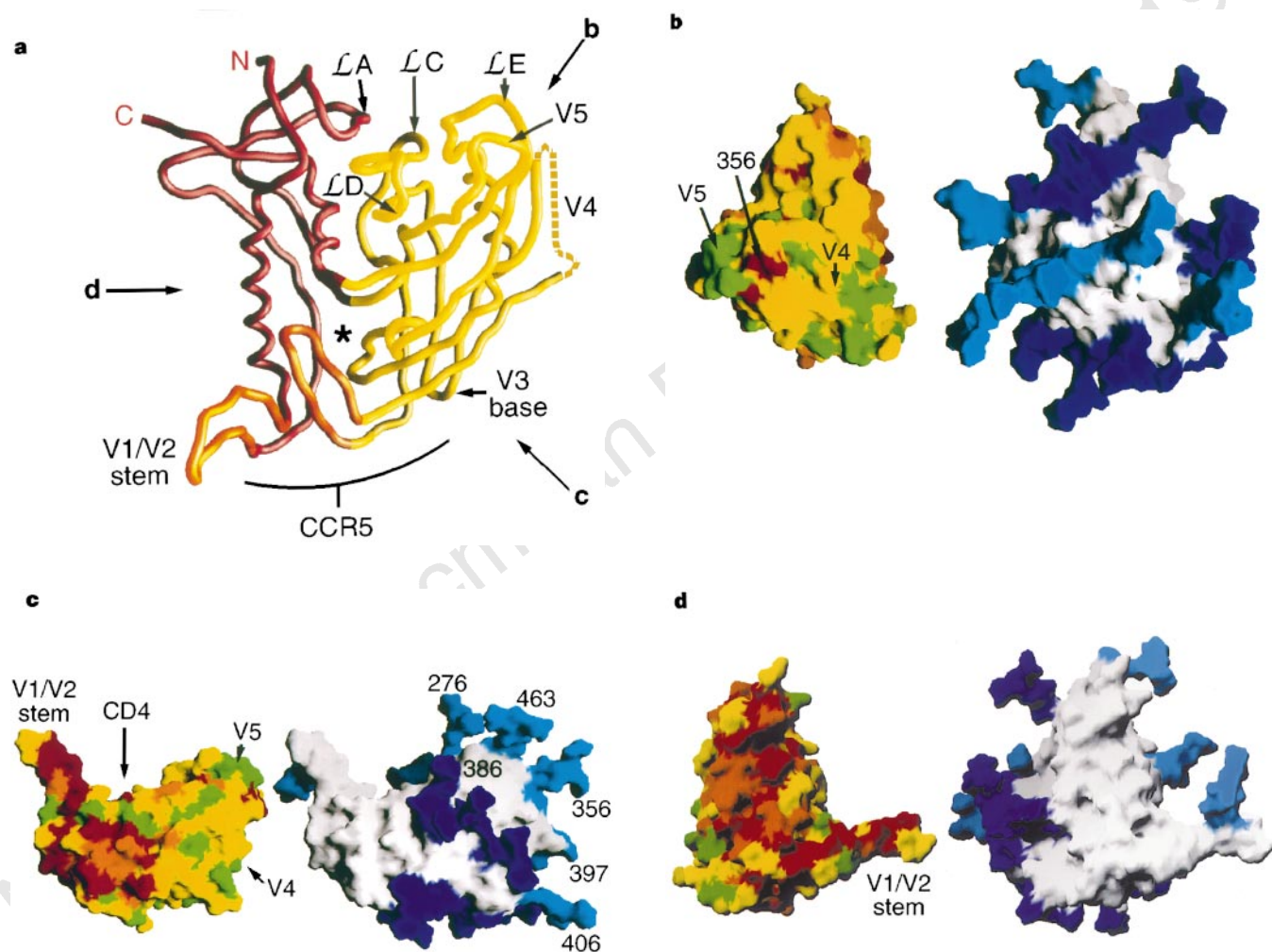


Figure 1 Structure and orientation of the HIV-1 gp120 core. **a**, A α tracing of the gp120 core, which was crystallized in a ternary complex with two-domain soluble CD4 and the Fab fragment of the 17b antibody¹², is shown. The gp120 core is seen from the perspective of CD4, and is oriented with the viral membrane at the top of the figure and the target cell membrane at the bottom. The inner gp120 domain is shown in red and the outer domain in yellow; the 'bridging sheet' is orange. The N and C termini of the truncated gp120 core are labelled, as are the positions of structures related to the gp120 variable regions V1–V5. The $\angle$ A, $\angle$ C, $\angle$ D and $\angle$ E surface loops¹² are shown. The position of the Phe43 cavity involved in CD4 binding is indicated by an asterisk. The gp120 surface implicated in binding to the CCR5 chemokine receptor (C. Rizzuto and J.G.S., submitted) is shown. The perspectives shown in **b–d** are indicated. **b**, View of the molecular surface of the gp120 outer domain, from the perspective indicated in **a**. The molecular surface on the left is coloured according to the variability observed in gp120 residues among primate immunodeficiency viruses: red, residues conserved among all primate immunodeficiency viruses; orange, residues conserved in all HIV-1 isolates; yellow, residues exhibiting some variation among HIV-1 isolates; and green, residues showing significant variability among HIV-1 isolates (see Methods). The variability of the gp120 surface is underestimated here because the V4 variable loop, which is not resolved in the structure, contributes to this

surface (approximate location is indicated). The position of the V5 region is shown. Also note the highly conserved glycosylation site (Asn356 and Thr/Ser358) within the $\angle$ E loop, between the V5 and V4 regions. On the right, the V4 loop and the carbohydrates are modelled (see Methods). The complex carbohydrate addition sites used in mammalian cells¹⁴ are coloured light blue, and the high-mannose sites are dark blue. The gp120 protein surface is white. **c**, View of the gp120 molecular surface that faces the target cell. Variability is indicated on the left, using the same colour scheme as in **b**. Note the clear demarcation between the conserved surface, which has been implicated in the formation of CD4i epitopes¹⁸ and in chemokine-receptor binding (C. Rizzuto and J.G.S., submitted), and the variable surface of the outer domain. The recessed binding site for CD4 is indicated, flanked by the V1/V2 stem, which is labelled. The V4 loop and the carbohydrates are modelled on the right (colouring as in **b**). Particular carbohydrates referred to in the text are labelled. **d**, View of the molecular surface of the gp120 core inner domain. Variability is indicated on the left by the colour scheme used in **b**. The CD4-binding site is on the right; the protruding V1/V2 stem is indicated. The conserved molecular surface, which is associated with the inner domain of the gp120 core, is devoid of known *N*-linked glycosylation sites. These are modelled on the right, which is coloured as in **b**.

tion by antibodies against V3 and CD4i epitopes²³. The latter effect is mediated primarily by the V2 loop²¹, suggesting that part of the V2 loop folds back along the V1/V2 stem to mask the 'bridging sheet' and adjacent V3 loop. The proximity of the V2 and V3 loops is supported by the observation that, in monkeys infected with simian-human immunodeficiency viruses (SHIVs), neutralizing antibodies are raised against discontinuous epitopes with V2 and V3 components (B. Etemad-Moghadam *et al.*, submitted). The CD4i epitopes are probably masked by the flanking V2 and V3 loops,

requiring the evolution of antibodies with protruding ('male') complementarity-determining regions (CDRs) to access these conserved epitopes. It has been suggested that CD4 binding repositions the V1/V2 loops, thus exposing the CD4i epitopes²¹. The presence of contacts between the V1/V2 stem and CD4 in the crystal structure¹² is consistent with this model.

CD4BS epitopes. CD4 makes several contacts within a recessed pocket on the gp120 surface. The gp120-CD4 interface includes two cavities, one water-filled and bounded equally by both proteins,

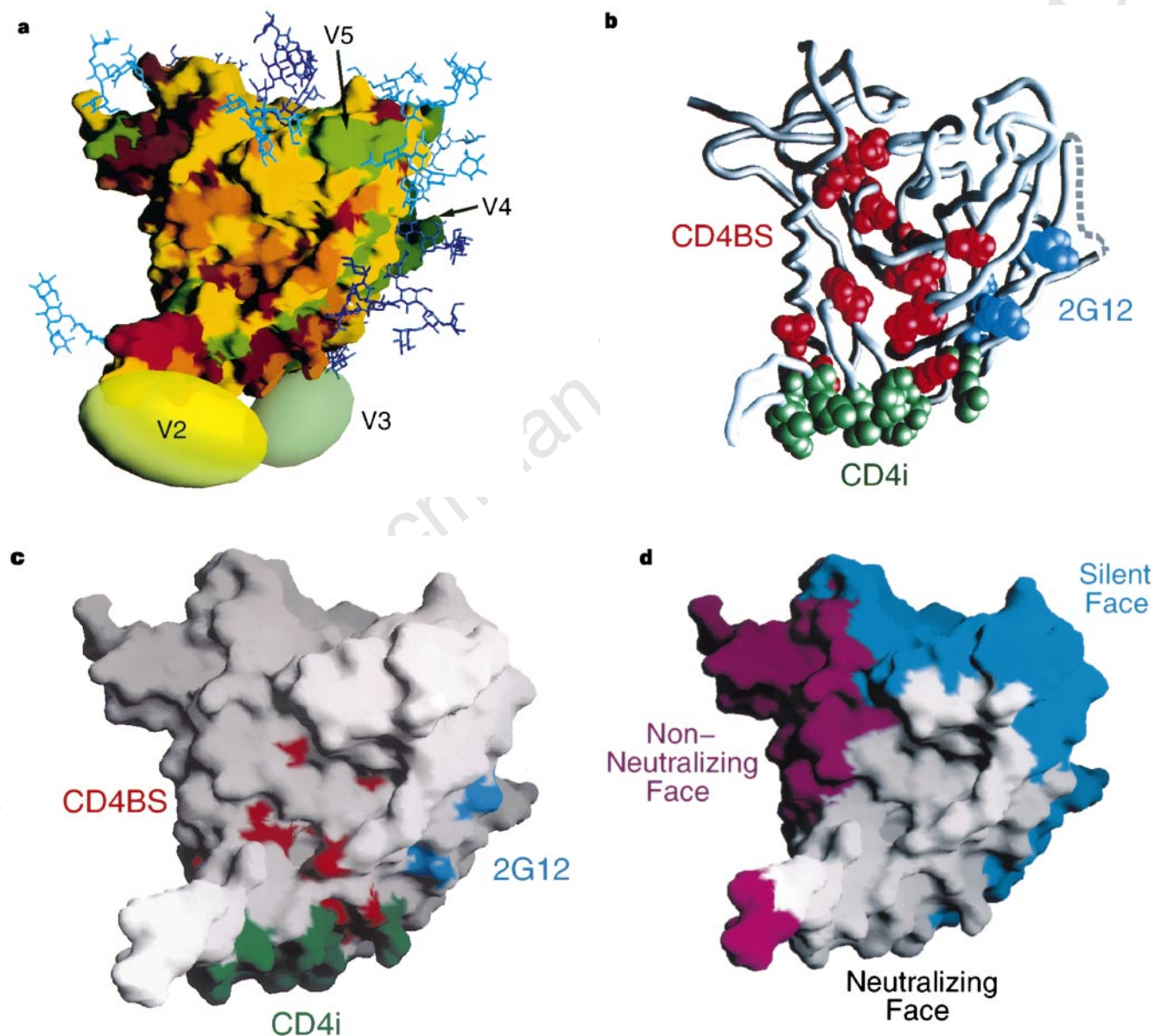


Figure 2 The spatial relationship of epitopes on the HIV-1 gp120 glycoprotein. **a**, The molecular surface of the gp120 core is shown, using the same perspective as in Fig. 1a. The modelled N-terminal gp120 core residues, V4 loop and carbohydrate structures are included. The variability of the molecular surface is indicated (colour scheme as in Fig. 1b). The modelled carbohydrates are shown in light blue (complex sugars) or dark blue (high-mannose sugars). The approximate locations of the V2 and V3 variable loops are indicated. Note the well-conserved surfaces near the 'Phe43' cavity and the chemokine-receptor-binding site (Fig. 1a). **b**, Cα tracing of the gp120 core, oriented as in Fig. 1a. The gp120 residues within 4 Å of the 17b CD4i antibody are shown in green; those implicated in the binding of CD4BS antibodies²² are in red. Changes in these residues significantly affect the binding of at least 25% of the CD4BS antibodies listed in Table 1. The

residues implicated in 2G12 binding¹⁹ are shown in blue. The V4 variable loop, which contributes to the 2G12 epitope¹⁹, is indicated by dotted lines. **c**, The molecular surface of the gp120 core, oriented and coloured as in **b**. **d**, Approximate locations of the faces of the gp120 core, defined by the interaction of gp120 and antibodies. The molecular surface accessible to neutralizing ligands (CD4 and CD4BS, CD4i and 2G12 antibodies) is shown in white. The neutralizing face of the complete gp120 glycoprotein includes the V2 and V3 loops, which are found adjacent to the surface shown (**a**). The approximate location of the gp120 face that is poorly accessible on the assembled envelope glycoprotein trimer and therefore elicits only non-neutralizing antibodies^{5,6} is shown in magenta. The approximate location of an immunologically 'silent' face of gp120, which roughly corresponds to the highly glycosylated outer domain surface, is in blue.

the other extending into the gp120 interior and contacting CD4 only at Phe 43 (Fig. 1a)¹². Table 1 and Fig. 2b, c show the gp120 residues implicated in the formation of CD4BS epitopes recognized by eight representative antibodies. CD4BS epitopes are uniformly disrupted by changes in Asp 368 and Glu 370 (ref. 22), which surround the opening of the 'Phe 43 cavity'. These residues are located on a ridge at the intersection of the two receptor-binding gp120 surfaces, consistent with competition studies suggesting that CD4BS epitopes overlap both the CD4i epitopes and the binding site for CD4 (refs 5, 18). The location of the gp120 residues implicated in formation of the CD4BS epitopes suggests that important elements of the CD4-binding surface of gp120 are accessible to antibodies.

Some CD4BS antibodies, like IgG1b12, are particularly potent at neutralizing HIV-1 (ref. 24). IgG1b12 binding is disrupted by gp120 changes that affect the binding of other CD4BS antibodies but, atypically, is sensitive to changes in the V1/V2 stem-loop structure²⁵. The observation that some well-conserved residues in the gp120 V1/V2 stem contact CD4 (ref. 12) raises the possibility that this protruding structure also contributes to the IgG1b12 epitope. This might increase the ability of the antibody to access the assembled envelope glycoprotein trimer, thus increasing neutralizing capability.

Although the CD4BS epitopes and the CD4-binding site overlap, several observations demonstrate that the binding of CD4BS antibodies differs from that of CD4. Changes in Trp 427, a gp120 residue that contacts both the Phe 43 cavity and CD4, uniformly disrupt CD4 binding but affect the binding of only some CD4BS antibodies (Table 1). Conversely, some changes in other cavity-lining gp120 residues, Ser 256 and Thr 257, affect the binding of CD4BS antibodies more than the binding of CD4 (ref. 22). As the recessed position of Ser 256 and Thr 257 in the current crystal structure (Fig. 2b, c) makes direct contacts with antibody unlikely, either the effects of changes in these residues are indirect or the CD4BS antibodies recognize a gp120 conformation that differs from the CD4-bound state. With respect to the latter possibility, it is interesting that several of the residues implicated in the integrity of the CD4BS epitopes are located in the interface between the inner and outer gp120 domains. CD4BS antibodies might recognize a gp120 conformation in which the spatial relationship between the domains is altered compared with the CD4-bound state, thus allowing better surface exposure of these residues. Differences between the CD4BS epitopes and the CD4-binding site create opportunities for neutralization escape²². The gp120 residues surrounding the Phe 43 cavity are highly conserved among primate immunodeficiency viruses (Fig. 2a), but the observed modest variation in adjacent surface-accessible residues (for example, Pro 369, Thr 373 and Lys 432) could account for decreased recognition of the gp120 glycoprotein from some geographic clades of HIV-1 by CD4BS antibodies²⁵. Additional potential for variation near or within the CD4BS epitopes is created by the unusual water-filled cavity in the gp120-CD4 binding interface, because CD4 binding can apparently tolerate change in the gp120 residues contacting this cavity¹².

The recessed nature of the CD4-binding pocket on gp120 (Fig. 1c) may delay the generation of high-affinity antibodies against the CD4BS epitopes and may afford opportunities to minimize the antiviral efficacy of such antibodies once they are elicited. The degree of recession is probably much greater on the full-length glycosylated gp120 than is evident on the crystallized gp120 core. The recessed pocket is flanked on one side by the V1/V2 stem-loop structure. The characterization of HIV-1 escape mutants from the IgG1b12 CD4BS antibody and the mapping of several V2 conformational epitopes support a model in which the V2 loop folds back along the V1/V2 stem, with V2 residues 183-188 proximal to Asp 368 and Glu 370. This model is consistent with observations that V1/V2 changes, in combination with V3 changes, can alter the

exposure of the adjacent CD4BS epitopes, particularly on the assembled trimer²⁶. The high temperature factors associated with the V1/V2 stem¹² imply flexibility in this protruding element (Fig. 1c, d), expanding the potential range of space occupied by the V1/V2 stem-loop structure. This could increase masking of the adjacent CD4BS and CD4i gp120 epitopes and divert antibody responses towards the variable loops.

Glycosylation may modify the interaction of antibodies with CD4BS epitopes. The LD loop, on the rim of the CD4-binding pocket opposite the V1/V2 stem, contains a well-conserved glycosylation site, Asn 276 (Fig. 1c). Changes in this site and at the adjacent Ala 281 have been associated with escape from the neutralizing activity of patient sera²⁷ and have been seen in SHIVs extensively passaged in monkeys²⁸. Another conserved glycosylation site at Asn 386 lies adjacent to both CD4BS and CD4i epitopes (Fig. 1c) and could diminish antibody responses against those sites. Additionally, in various HIV-1 strains, carbohydrates are added to the V2 loop segment (residues 186-188) thought to be proximal to the CD4BS epitopes.

The 2G12 epitope. The integrity of the 2G12 epitope is disrupted by changes in gp120 glycosylation, by either glycosidase treatment or mutagenic alteration of specific N-linked carbohydrate-addition sites¹⁹. These sites are located on the relatively variable surface of the gp120 outer domain, opposite to and approximately 25 Å away from the CD4-binding site (Fig. 2b, c). The gp120 glycoprotein synthesized in mammalian cells exhibits a dense concentration of high-mannose sugars in this region (Fig. 2a). Even in the enzymatically deglycosylated gp120 core, carbohydrate residues constitute much of this surface. 2G12 probably binds at least in part to these

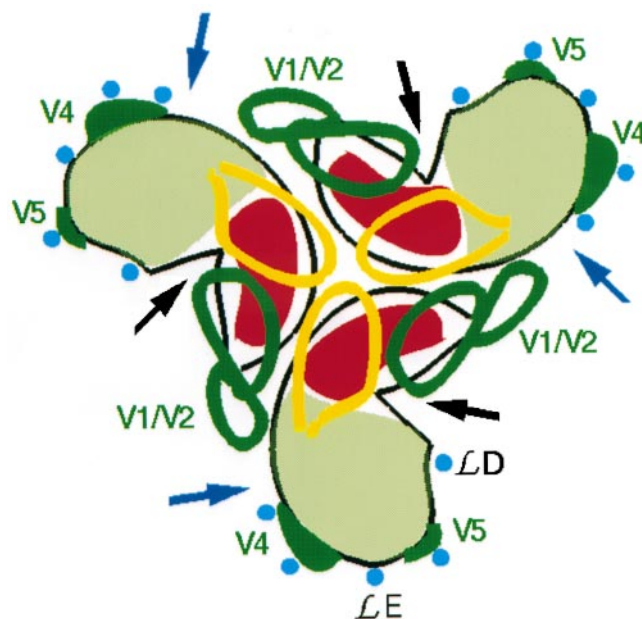


Figure 3 A likely arrangement of the HIV-1 gp120 glycoproteins in a trimeric complex. The gp120 core was organized into a trimeric array, based on the criteria discussed in the text. The perspective is from the target-cell membrane, like that in Fig. 1c. The CD4-binding pockets are indicated by black arrows, and the conserved chemokine-receptor-binding regions are in red. Areas shaded in light green indicate the more variable, glycosylated surfaces of the gp120 cores. The approximate locations of the 2G12 epitopes are indicated by blue arrows; those of the V3 loops (yellow) and V4 regions (green) are indicated. The positions of the V5 regions (green) and some complex-carbohydrate addition sites (asparagines 276, 463, 356, 397 and 406) (blue dots) are shown. The approximate locations of the large V1/V2 loops, centred on the known positions of the V1/V2 stems, are indicated (green). On one of the gp120 subunits, the positions of the LD and LE loops are indicated. The distance of each of the gp120 monomers from the 3-fold symmetry axis is arbitrary.

carbohydrates, explaining the surprising conservation of the 2G12 epitope despite the variability of the underlying protein surface, which includes the stem of the V3 loop and the V4 variable region. The inclusion of carbohydrate in the epitope might also explain the apparent rarity with which these antibodies are generated. The localization of the 2G12 epitope is consistent with previous studies indicating that 2G12 forms a unique competition group^{5,19} and does not interfere with the binding of monomeric gp120 to either CD4 or chemokine receptors¹¹. As the 2G12 epitope is predicted to be oriented towards the target cell upon CD4 binding (see below), the antibody may sterically impair interactions of the oligomeric envelope glycoprotein complex with host cell moieties.

Possible orientations of the exterior glycoproteins in the trimer are significantly constrained by the requirement that observed and deduced binding sites for receptors and neutralizing antibodies, sites of *N*-linked glycosylation, and variable structures be exposed on the surface of the assembled complex. The two-domain CD4 in the ternary complex structure was aligned to the structure of four-domain CD4 (ref. 29) to orient the trimer model with respect to the target cell membrane. The predictions of such a model (Fig. 3) are: (1) the chemokine-receptor-binding sites are clustered at the vertex of the trimer predicted to be closest to the target cell; (2) both variable and conserved neutralization epitopes are concentrated on the half of gp120 facing the target cell; (3) possibilities for inter-subunit interactions among the variable structures that could help mask conserved neutralization epitopes are created; (4) the subset of gp120 glycosylation sites to which complex carbohydrates are added in mammalian cells¹⁴ is well exposed on the outer periphery of the trimer; (5) the highly conserved surface near the α 1 helix is available for gp41 and/or gp120 protein interactions within the trimers; and (6) the surface of the assembled envelope glycoprotein complex is roughly hemispherical, minimizing the surface area of the viral spike that is potentially exposed to antibodies.

In summary, the X-ray crystal structure of the gp120 core/two-domain CD4/17b Fab complex provides a framework for visualizing important interactions between HIV-1 and the humoral immune system. Previous antibody competition analyses indicated that the gp120 surface buried in the assembled trimer elicits non-neutralizing antibodies^{5,6}. By contrast, the binding sites for neutralizing antibodies cluster on a different gp120 surface⁵. Our structural studies support the existence of non-neutralizing and neutralizing faces of gp120, and reveal another, immunologically 'silent' face of the glycoprotein (Fig. 2d). This outer domain surface, together with the principal variable loops, contributes to the large fraction of the gp120 surface that is protected against antibody responses by a dense array of carbohydrates and by the capacity for variation. The conserved receptor-binding regions of gp120 represent attractive targets for immune intervention. However, the elicitation of antibodies against these conformation-dependent structures is inefficient. As the gp120 epitopes near the receptor-binding regions span the inner and outer domains, interdomain conformational shifts may decrease their representation in the immunogen pool. The recessed nature of the CD4-binding site probably contributes to its poor immunogenicity. The sequential recognition of two receptors by primate immunodeficiency viruses allows the conserved elements of the chemokine-receptor-binding site to be created or exposed only after CD4 binding has occurred. At that point, it is likely that the proximity of the chemokine-receptor-binding site to the cell membrane sterically limits antibody binding. The evolution of primate immunodeficiency viruses that persist successfully, despite the host immune response, presents a challenge to vaccine development. An understanding of the structures of the relevant gp120 epitopes should assist efforts to overcome these hurdles. □

Methods

Graphics. Molecular graphics were produced using Midas-Plus (University of California, San Francisco) and GRASP³⁰.

Assignment of variability. Variability in gp120 residues was assessed using an alignment of sequences derived from ~400 HIV-1, HIV-2 and simian immunodeficiency viruses¹³. Residues were assigned variability indices and colour coded as follows: red, conserved in all primate immunodeficiency viruses; orange, conserved in all HIV-1, including groups M and O and chimpanzee isolates; yellow, some variation among HIV-1 isolates (divergence from the consensus sequence in 1–8 of the 12 HIV-1 groups examined); green, variable among HIV-1 isolates (divergence from the consensus sequence in ≥ 9 of the 12 HIV-1 groups examined).

Molecular modelling. Residues 88, 89 and 397–409, which are disordered in the ternary complex crystals¹², were built manually using the program TOM. For the V4 loop (residues 397–409), a dominant constraint was the distance between the ordered residues 396 and 410 (C α –C α distance of 26.88 Å). For the carbohydrate, examination of the *N*-linked carbohydrate in several crystal structures (for example, 1fc2, 1gly, 1lte) showed that the core common to both high-mannose and complex *N*-linked sugars, (NAG)₂(MAN)₃, did not differ greatly in conformation after alignment of the first NAG (*N*-acetylglucosamine) (MAN, mannose). This core, which represents roughly half the total glycosylation for a typical *N*-linked site, was built onto each of the 18 consensus *N*-linked glycosylation sites found on the HXBc2 gp120 core. The stereochemistry of this initial model was refined using simulated annealing in XPLOR. Briefly, the model was heated to temperatures of between 2,500 and 3,500 K, and slow cooled in steps of 25 to 300 K. At each step, molecular dynamics were performed with the core gp120 fixed, allowing only the modelled residues and carbohydrate (including any attached Asn) to move. In three separate runs, performing molecular dynamics for 5 fs per step, all steric clashes could be removed and the geometry idealized, with an average root-mean-square (r.m.s.) of carbohydrate movement of only ~3.5 Å. Four subsequent runs were made using dynamic times of between 50–75 fs per step. The carbohydrate positions obtained from these runs differed substantially from those in the starting model (average carbohydrate r.m.s. difference of ~8 Å). Two of the models from these longer annealings were more similar to each other than to the rest (r.m.s. differences in carbohydrate of ~4 Å versus ~8 Å for all other models): one had been heated to 3,500 K, with dynamics of 75 fs per step; the other (see figures) was heated to only 2,500 K, with dynamics of 50 fs per step. In general, the r.m.s. movement of the NAG sugars was roughly half the r.m.s. movement of the MAN sugars, reflecting greater conformational flexibility away from the protein surface.

Received 15 May; accepted 27 May 1998.

- Allen, J. *et al.* Identification of the major envelope glycoprotein product of HTLV-III. *Science* **228**, 1091–1094 (1983).
- Dalgleish, A. G. *et al.* The CD4 (T4) antigen is essential component of the receptor for the AIDS retrovirus. *Nature* **312**, 763–767 (1984).
- Klatzmann, D. *et al.* T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV. *Nature* **312**, 767–768 (1984).
- Feng, Y., Broder, C. C., Kennedy, P. E. & Berger, E. HIV-1 entry cofactor: Functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science* **272**, 872–877 (1996).
- Moore, J. P. & Sodroski, J. Antibody cross-competition analysis of the human immunodeficiency virus type 1 gp120 exterior envelope glycoprotein. *J. Virol.* **70**, 1863–1872 (1996).
- Wyatt, R. *et al.* Analysis of the interaction of the human immunodeficiency virus type 1 (HIV-1) gp120 envelope glycoprotein with the gp41 transmembrane glycoprotein. *J. Virol.* **71**, 9722–9731 (1997).
- Sattentau, Q. J. & Moore, J. P. Human immunodeficiency virus type 1 neutralization is determined by epitope exposure on the gp120 oligomer. *J. Exp. Med.* **182**, 185–196 (1995).
- Posner, M. *et al.* An IgG human monoclonal antibody which reacts with HIV-1 gp120, inhibits virus binding to cells, and neutralizes infection. *J. Immunol.* **146**, 4325–4332 (1991).
- Ho, D. *et al.* Conformational epitope on gp120 important in CD4 binding and human immunodeficiency virus type 1 neutralization identified by a human monoclonal antibody. *J. Virol.* **65**, 489–493 (1991).
- Wu, L. *et al.* CD4-induced interaction of primary HIV-1 gp120 glycoproteins with the chemokine receptor CR5. *Nature* **384**, 179–183 (1996).
- Trkola, A. *et al.* CD4-dependent, antibody-sensitive interactions between HIV-1 and its co-receptor CCR-5. *Nature* **384**, 184–187 (1996).
- Kwong, P. D. *et al.* Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* **393**, 648–659 (1998).
- Myers, G. *et al.* Human Retroviruses and AIDS. A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences (Los Alamos National Laboratory, 1996).
- Leonard, C. *et al.* Assignment of interchain disulfide bonds and characterization of potential glycosylation sites of the type 1 human immunodeficiency virus envelope glycoprotein (gp120) expressed in Chinese hamster ovary cells. *J. Biol. Chem.* **265**, 10373–10382 (1990).
- Fung, M. S. C. *et al.* Identification and characterization of a neutralization site within the second variable region of human immunodeficiency virus type 1 gp120. *J. Virol.* **66**, 848–856 (1992).
- Putney, S. *et al.* HTLV-III/LAV-neutralizing antibodies to an *E. coli*-produced fragment of the virus envelope. *Science* **234**, 1392–1395 (1986).
- Rusche, J. R. *et al.* Antibodies that inhibit fusion of human immunodeficiency virus-infected cells bind a 24-amino-acid sequence of the viral envelope gp120. *Proc. Natl Acad. Sci. USA* **85**, 3198–3202 (1988).
- Thali, M. *et al.* Characterization of conserved human immunodeficiency virus type 1 gp120 neutralization epitopes exposed upon gp120-CD4 binding. *J. Virol.* **67**, 3978–3988 (1993).

19. Trkola, A. *et al.* Human monoclonal antibody 2G12 defines a distinctive neutralization epitope on the gp120 glycoprotein of human immunodeficiency virus type 1. *J. Virol.* **70**, 1100–1108 (1996).
20. Binley, J. *et al.* Analysis of the interaction of antibodies with a conserved, enzymatically deglycosylated core of the HIV-1 gp120 envelope glycoprotein. *AIDS Res. Hum. Retrovir.* **14**, 191–198 (1997).
21. Wyatt, R. *et al.* Involvement of the V1/V2 variable loop structure in the exposure of human immunodeficiency virus type 1 gp120 epitopes induced by receptor binding. *J. Virol.* **69**, 5723–5733 (1995).
22. Thali, M. *et al.* Discontinuous, conserved neutralization epitopes overlapping the CD4 binding region of the HIV-1 gp120 envelope glycoprotein. *J. Virol.* **66**, 5635–5641 (1992).
23. Cao, J. *et al.* Replication and neutralization of human immunodeficiency virus type 1 lacking the V1/V2 variable loops of the gp120 envelope glycoprotein. *J. Virol.* **71**, 9808–9812 (1997).
24. Roben, P. *et al.* Recognition properties of a panel of human recombinant Fab fragments to the CD4 binding site of gp120 that show differing abilities to neutralize human immunodeficiency virus type 1. *J. Virol.* **68**, 4821–4848 (1994).
25. Moore, J. P. *et al.* Exploration of antigenic variation in gp120 from clades A through F of human immunodeficiency virus type 1 by using monoclonal antibodies. *J. Virol.* **68**, 8350–8364 (1994).
26. Wyatt, R. *et al.* Functional and immunologic characterization of human immunodeficiency virus type 1 envelope glycoproteins containing deletions of the major variable regions. *J. Virol.* **67**, 4557–4565 (1993).
27. Watkins, B. A. *et al.* Immune escape by human immunodeficiency virus type 1 from neutralizing antibodies: evidence for multiple pathways. *J. Virol.* **67**, 7493–7500 (1993).
28. Karlsson, G. *et al.* Characterization of molecularly cloned simian-human immunodeficiency viruses causing rapid CD4⁺ lymphocyte depletion in rhesus monkeys. *J. Virol.* **71**, 4218–4225 (1997).
29. Wu, H., Kwong, P. D. & Hendrickson, W. A. Dimeric association and segmental variability in the structure of human CD4. *Nature* **387**, 527–530 (1997).
30. Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281–296 (1991).

Acknowledgements. We thank Y. McLaughlin and S. Farnum for manuscript preparation and M. Farzan for discussion. This research was supported by the NIH, the Howard Hughes Medical Institute, the American Foundation for AIDS Research and the Aaron Diamond Foundation. This work was made possible by gifts from the late William McCarty-Cooper, from the G. Harold and Leila Y. Mathers Foundation, from the Friends 10, and from Douglas and Judi Krupp.

Correspondence and requests for materials should be addressed to J.G.S.

Dynein arms are oscillating force generators

Chikako Shingyoji, Hideo Higuchi*†, Misako Yoshimura, Eisaku Katayama‡ & Toshio Yanagida*§

Department of Biological Sciences, Graduate School of Science, University of Tokyo, Hongo, Tokyo 113-0033, Japan

* Yanagida BioMotron Project, ERATO, JRDC, Senba-Higashi 2-4-14, Mino, Osaka 562-0035, Japan

‡ Department of Fine Morphology, Institute of Medical Science, University of Tokyo, Minato-ku, Tokyo 108-0071, Japan

§ Department of Physiology, Osaka University Medical School, Suita, Osaka 565-0871, Japan

Eukaryotic flagella beat rhythmically¹. Dynein is a protein that powers flagellar motion, and oscillation may be inherent to this protein^{2–5}. Here we determine whether oscillation is a property of dynein arms themselves or whether oscillation requires an intact axoneme⁶, which is the central core of the flagellum and consists of a regular array of microtubules. Using optical trapping nanometry^{7,8}, we measured the force generated by a few dynein arms on an isolated doublet microtubule. When the dynein arms on the doublet microtubule contact a singlet microtubule and are activated by photolysis of caged ATP⁸, they generate a peak force of ~6 pN and move the singlet microtubule over the doublet microtubule in a processive manner. The force and displacement oscillate with a peak-to-peak force and amplitude of ~2 pN and ~30 nm, respectively. The geometry of the interaction indicates that very few (possibly one) dynein arms are needed to generate the oscillation. The maximum frequency of the oscillation at 0.75 mM ATP is ~70 Hz; this frequency decreases as the ATP concentration decreases. A similar oscillatory force is also generated by inner dynein arms alone on doublet microtubules that are depleted of outer dynein arms. The oscillation of the dynein arm may be a basic mechanism underlying flagellar beating.

A crucial step in our study was the preparation of functionally intact dynein arms. ATP was photoreleased to activate demembrated and elastase-treated⁴ sea-urchin sperm flagella. This activation caused disintegration of the axoneme into individual doublet microtubules. This procedure exposed fresh, functional dynein arms⁹ on the doublets (Fig. 1a). When a suspension of singlet microtubules was added, the singlet microtubules attached themselves to the doublets and, after flash photolysis of caged ATP, moved along the doublets. Figure 1c shows two microtubules attached to the dynein arms of a doublet. Using polarity marked microtubules, we confirmed that the sliding movement was always towards the plus (fast-growing) end of the microtubule. This is consistent with the known polarity of force generation by dynein^{9–12}, indicating that the movement was caused by the dynein arms. The maximum velocity, obtained ~100 ms after flash photolysis, was 6.4 $\mu\text{m s}^{-1}$ (see below). This is comparable to

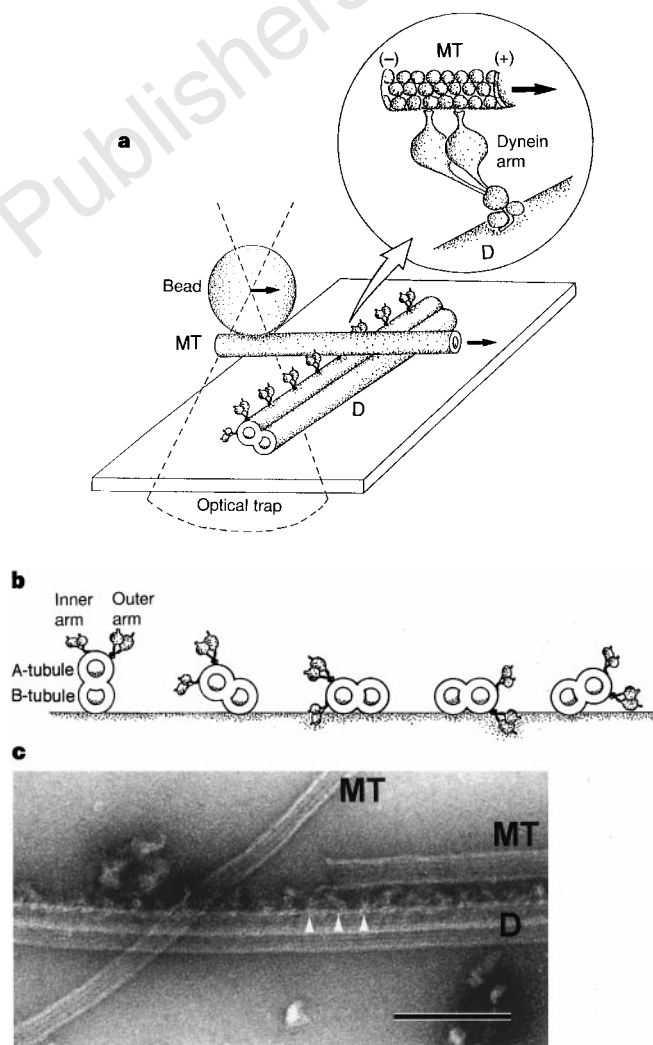


Figure 1 Interactions of singlet microtubules with dynein arms present on doublet microtubules. **a**, Schematic diagram of exposed dynein arms, on a doublet microtubule (D), interacting with a singlet microtubule (MT). The singlet microtubule is manipulated by means of an optically trapped streptavidine-coated bead. A two-headed arm⁹ pulls the singlet microtubule in the direction of its plus (+) end, causing the bead to move away from the centre of the trap force (black arrows). **b**, Possible orientations of doublet microtubules on the glass surface (bottom) with outer and inner dynein arms. Dynein arms pointing upwards may be capable of interaction with a singlet microtubule (not shown). **c**, Electron micrograph showing two singlet microtubules (MT) interacting with dynein arms (white arrowheads) on a doublet microtubule (D). Scale bar, 100 nm.

† Present address: Department of Metallurgy, Faculty of Engineering, Tohoku University, Aramaki-aza-Aoba, Sendai 980-8579, Japan.

the velocity of microtubules translocated by dissociated dynein ($4\text{--}8\ \mu\text{m s}^{-1}$)^{12,13}.

To manipulate the microtubule, we applied optical forces to streptavidine-coated latex beads that had been attached as handles to biotinylated single microtubules (Fig. 1a). When the free end of such a microtubule was brought into contact with the surface of a doublet, the microtubule stopped exhibiting Brownian motion and, after photolysis of caged ATP, began to move. If the angle between the singlet microtubule and the doublet was less than $\sim 30^\circ$, the singlet microtubule gradually became parallel to the doublet and moved along the doublet until finally the bead escaped from the optical trap force (at forces of $>30\text{--}40\text{ pN}$). However, when the microtubule was placed obliquely (at $45\text{--}90^\circ$) on the doublet (Fig. 1a), the bead did not escape and force was recorded repeatedly after each flash of ultraviolet light.

Figure 2a shows a typical record of the force generation. Following a rising phase, in which the microtubule was displaced at $5.3\ \mu\text{m s}^{-1}$, the force was maintained for periods of between several hundred milliseconds and several seconds. In Fig. 2b we summarize the maintained forces. The histogram shows a peak at about 6 pN. For doublet microtubules from which the outer dynein arms had been removed, a force similar to that shown in Fig. 2a was also generated by probably a single inner dynein arm (see below); the maintained force was $5.6 \pm 3.2\text{ pN}$ (mean $\pm$ s.d.; $n = 10$) and the rate of displacement in the rising phase was $2.1 \pm 0.8\ \mu\text{m s}^{-1}$ ($n = 7$). The slower rate of displacement of singlet microtubules from outer-arm-depleted doublet microtubules may be related to the lower sliding velocity of the inner dynein arms than of the outer arms^{11–14}. Figure 2a also shows that the dynein displaced the singlet

microtubule by up to 100 nm without detaching from it. This contradicts the idea that an axonemal dynein may not be capable of processive movement along a microtubule¹⁵. Kinesin is a processive motor^{7,16}, whereas myosin is not^{17,18}. Our results indicate that dynein is a processive motor and, in this respect, resembles kinesin.

We estimated the number of dynein arms involved in the observed force generation as follows. As the diameter of a microtubule (25 nm) is comparable to the axial period of the dynein arms on the doublet (24 nm)^{19,20}, a singlet microtubule crossing a doublet at an angle of $45\text{--}90^\circ$ could interact either with a single dynein arm or with very few (2–4) arms. The microtubule could interact with both outer and inner arms only when the doublet stood 'on edge', with the B-tubule of the doublet downwards (Fig. 1b, left). In this configuration, the microtubule can interact with 2–4 dynein arms. In other cases, however, only one or two of either the outer or the inner arms can interact with the microtubule (Fig. 1b, c). To confirm the number of dynein arms involved in the force generation, we used outer-arm-depleted doublet microtubules. In the outer-arm-depleted doublets, only one or at most two of the inner dynein arms can interact with the microtubule. The maximum maintained forces observed at about 90° , 60° and 45° were 15.8 pN, 20.6 pN and 26.6 pN, respectively (Fig. 2b). Thus, the maximum number of dynein arms that could possibly interact with the microtubule increases with a decrease in the angle of interaction. On the other hand, the peak force was always $\sim 6\text{ pN}$ (Fig. 2b), indicating that the unit force generated at $\sim 90^\circ$ is not less than that at $\sim 60^\circ$ and $\sim 45^\circ$. These results, and the $\sim 6\text{ pN}$ peak in both the intact (Fig. 2b) and the outer-arm-depleted doublets, indicate that the unitary force of a dynein arm may be $\sim 6\text{ pN}$. This unitary force

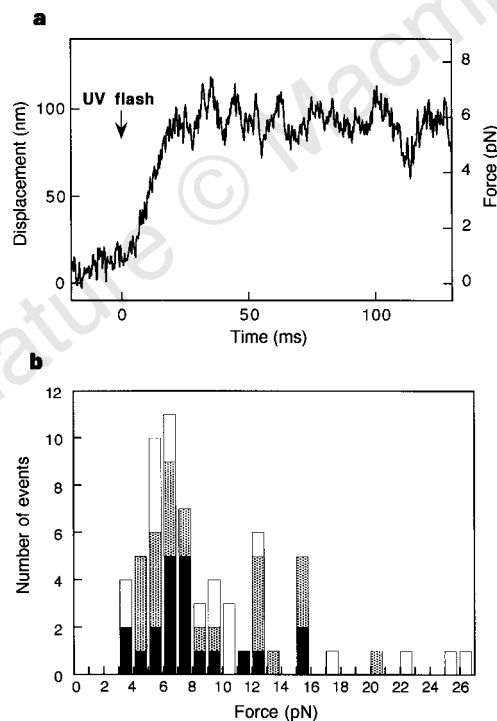


Figure 2 Force generation by dynein arms. **a**, Typical record of force generation. The angle between the singlet microtubule and the doublet microtubule to which the dynein arms were attached was 80° . An ultraviolet (UV) flash was applied at time zero (arrow). A low-pass filter with a cut-off frequency of 2,000 Hz was used. Displacement refers to the displacement of the singlet microtubule over the doublet by the dynein arms. **b**, Summary of the maximum forces developed ($n = 65$). Filled, dotted and open bars indicate angles of $\sim 90^\circ$, $\sim 60^\circ$ and $\sim 45^\circ$, respectively, between the singlet microtubule and the doublets.

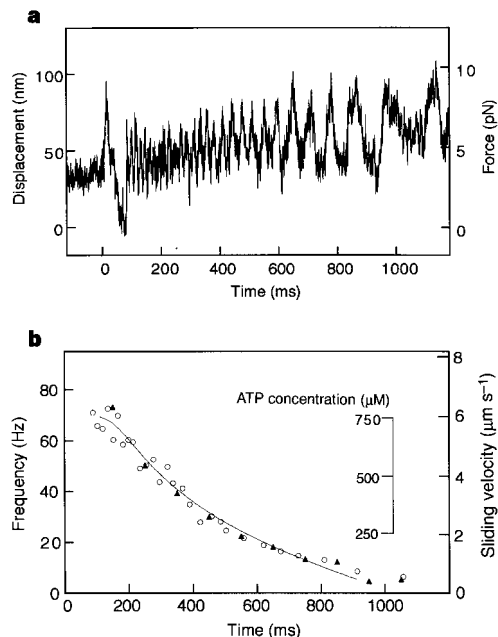


Figure 3 Force oscillations of dynein arms. **a**, Typical record of force oscillations. A flash of ultraviolet light was applied at time zero. The angle between the microtubule and the doublet was 70° . **b**, Changes in oscillation frequency obtained from **a** (open circles), and the velocities of sliding of untrapped microtubules along doublets determined in separate experiments (filled triangles), are shown. The line shows changes in the estimated ATP concentration.

is similar to the forces reported for single kinesin (5–7 pN)^{8,16} and single myosin (3–6 pN)^{17,18} molecules.

A remarkable feature of the dynein arm activity was the presence of oscillations (Figs 2a and 3a), which were observed in 33 of the 65 measurements used for Fig. 2b and 11 of the 14 measurements of the outer-arm-depleted doublets. The peak-to-peak forces and amplitudes of displacement of the oscillations were 2.3 ± 1.1 pN ($n = 20$) and 32.1 ± 10.7 nm ($n = 20$), respectively. The frequency was high at the beginning of the response and gradually decreased (Fig. 3a).

The decrease in frequency seems to be associated with the decrease in ATP concentration. The estimated concentration of the photoreleased ATP (see Methods) decreased in parallel with the frequency (Fig. 3b). The frequency of oscillations at 0.75 mM ATP was 72.2 ± 15.4 Hz ($n = 11$). Also, the velocities of untrapped microtubules that slid along the doublets decreased in parallel with the oscillation frequency (Fig. 3b). Thus, the frequency of the oscillations is a function of the ATP concentration.

The oscillation frequency reported here is close to the frequency of flagellar beating but this may be coincidental. It is much lower than the high-frequency vibration (~ 350 Hz) of the flagellar axoneme⁵. The power spectra of the oscillations in this study (Fig. 3a) obtained by a fast Fourier transform did not have a component of >100 Hz. Thus, the oscillations that we observed are distinct from the high-frequency vibration⁵, which may result from the activities of multiple dynein arms on multiple doublets within the axoneme.

Our results indicate that axonemal dynein acts processively. This may result from a hand-over-hand motion of the two heads of dynein^{11,19,20}, as has been predicted for kinesin⁷ and cytoplasmic dynein²¹, which have two homogeneous heads. The two heads of axonemal dynein of the sea-urchin sperm flagella, however, are heterogeneous^{14,20}. Such heterogeneity may be important in generating oscillation and in processive movement.

It is surprising that few, possibly only one, dynein arms on a single doublet microtubule can produce such oscillations. In an actomyosin system it is possible for a molecular motor under isotonic conditions to produce spontaneous self-sustained oscillations without having to couple to an external resonator, as occurs in the case of insect flight muscle^{22,23}. If we assume that a dynein head cycles through the three essential mechanical states of head attachment, tilting (force generation) and dissociation, with strain-dependent rate constants, and that linkage of a head to the doublet acts as a spring, then the dynein head may oscillate^{22,23}.

The dynein arm consists of two heterogeneous heads (heavy chains), which may behave differently in ATP hydrolysis and in force generation^{20,24,25}. Each heavy chain has four ATP-binding sites, although only one of these sites has been proved to be involved in ATP hydrolysis. At the ATP-hydrolysis step, several forms of dynein-product intermediate may be produced^{15,26}. Thus, it is possible that, during the mechanochemical cycle of a dynein arm, two dynein heads may undergo several chemical steps instead of one. If the dynein head, like kinesin, moves along the microtubule in 8-nm steps⁸ per head, using one ATP molecule per step, and if the backward motion of the oscillations also consists of ATP-driven steps, the number of ATP molecules needed for oscillations at 72 Hz with a 32-nm peak-to-peak amplitude would be 288 per second per dynein. This is of the same order of magnitude as the reported maximum ATP turnover rate of 150 ATP molecules per second per dynein²⁷, indicating that the oscillations could be explained as an ATP-coupled movement.

The dynein arm, a large protein⁹ of relative molecular mass (M_r) $\geq 1,200$ K, may behave as a functional unit consisting of multiple components^{9,19,20,28} that work together and are able to oscillate. Here, as the displacement of the singlet microtubule was increased to about 100 nm, the microtubule began to oscillate. Such behaviour is a kind of stretch activation²² of dynein, indicating a possible nonlinear²² property of the linkage between the dynein head and

the doublet. We have found that a dynein arm exhibits such high-order functions. These results may be important not only for understanding the mechanism of flagellar beating, but also for understanding the functions of motor proteins in general. □

Methods

Axonemes, microtubules and beads. Sperm of the sea urchin *Hemicentrotus pulcherrimus* were suspended in 2–3 volumes of Ca^{2+} -free artificial sea water² and demembrated with 500 volumes of extracting solution containing 150 mM K-acetate, 2 mM MgSO_4 , 20 mM HEPES, 2 mM EGTA, 0.04% (w/v) Triton X-100, 1 mM dithiothreitol (DTT), pH 7.8, for 45 s at 24–26°C. This suspension was diluted with 8 volumes of reactivating solution (150 mM K-acetate, 2 mM MgSO_4 , 20 mM HEPES, 2 mM EGTA, 2% (w/v) polyethyleneglycol (M_r 20K), 1 mM DTT, pH 7.8, without ATP), pelleted at 11,000 g and labelled with tetramethylrhodamine (25 μM , 4 min). The labelled sperm was pelleted at 17,000 g, resuspended in reactivating solution and homogenized to obtain axonemal fragments. Sperm heads were removed by centrifugation at 2,000 g; the supernatant containing axonemal fragments was centrifuged at 17,000 g and resuspended in reactivating solution. To remove outer arms, demembrated sperm were extracted with 0.6 M KCl (in reactivating solution without ATP) for 3 min before labelling with rhodamine. The beat frequency in 1 mM ATP of the outer-arm-depleted sperm was $\sim 48\%$ lower (13.3 ± 1.9 Hz, $n = 33$) than in the intact sperm (27.6 ± 2.7 Hz, $n = 50$), indicating $\geq 80\%$ extraction of the outer arms²⁹.

Microtubules were assembled by mixing bovine tubulin, tetramethylrhodamine-labelled tubulin⁸ and biotinylated tubulin with GTP and were stabilized with 40 μM taxol. Polarity-marked microtubules were constructed by polymerizing weakly rhodamine-labelled tubulin from the ends of strongly labelled tubulin seeds so that the plus ends grew to be much longer than the minus ends.

Carboxylated blue fluorescent beads, of 1 μm in diameter (Molecular Probes), were first crosslinked with biotin (0.1 mg ml^{-1}) using carbodiimide (5 mg ml^{-1}) and then coated with streptavidine (1 mg ml^{-1}).

Motility assay and optical trapping. We constructed a 15- μl perfusion chamber using two coverslips, with polyester sheets (Lumirror type 1, Toray Industries) as spacers. A suspension of axonemal fragments was introduced into the chamber and treated with elastase (5 $\mu\text{g ml}^{-1}$ elastase, Sigma type III, and 5 $\mu\text{g ml}^{-1}$ soybean trypsin inhibitor, Sigma type I-S, in reactivating solution) for 1–1.5 min. The elastase treatment was stopped with ovinhibitor (50 $\mu\text{g ml}^{-1}$ ovinhibitor, Sigma type IV-O in reactivating solution), and followed by a perfusion with casein (1 mg ml^{-1} in assay buffer) and a perfusion with assay buffer. A suspension of microtubules and beads in the caged ATP solution was then added. Assay buffer contained 70 mM K-acetate, 5 mM Mg-acetate, 20 mM HEPES, 0.1 mM EGTA, 0.1 mM EDTA and 1 mM DTT, pH 7.8. The caged ATP solution contained 1 mM caged ATP (P^3 -1-(2-nitrophenyl) ethyl ester of ATP, Dojin), 20 units per ml hexokinase (Toyobo), 10 μM taxol, 20 mM glucose, 0.5% (v/v) β -mercaptoethanol, 20 $\mu\text{g ml}^{-1}$ catalase and 100 $\mu\text{g ml}^{-1}$ glucose oxidase in assay buffer. After the perfusion, the spacer sheets were removed and the chamber was sealed with enamel.

The basic equipment used for disintegrating the axonemal fragments and for the force measurements was as described⁸. Fluorescent images of axonemal fragments, beads and microtubules were recorded with a high-sensitivity camera⁸. For the disintegration process and force measurement, ATP was photoreleased from caged ATP with an ultraviolet laser (λ 354 nm) for 10 ms. A streptavidine-coated bead was captured by the optical trap. Displacement of the bead by the activated arms was measured at a sampling rate of 4 kHz with a quadrant photodiode connected to a MacLab system (AD Instruments). All measurements were performed at 24–26°C.

The trap stiffness (0.04–0.14 pN nm^{-1}) and the bead-to-dynein linkage stiffness (0.1–0.3 pN nm^{-1}) were determined from the variance of the thermal fluctuation of beads¹⁶. Force was calculated as the displacement of the bead multiplied by the trap stiffness. The displacement of a microtubule was calculated from the bead displacements multiplied by 1.51 (for Fig. 2a) and 1.64 (for Fig. 3a), which are attenuation factors determined by the ratio of bead-to-dynein linkage stiffness to trap stiffness⁸. The changes in the concentration of photoreleased ATP with time have been estimated from the rate (22 molecules per second) of dark reaction of caged-ATP photolysis, the size of the

illumination area (20 μm in diameter), the diffusion constant of ATP (300 $\mu\text{m}^2 \text{s}^{-1}$) and the hexokinase activity (330 $\mu\text{M ATP s}^{-1}$)⁸.

Electron microscopy. The suspension of axonemal fragments prepared as above was applied onto one side of an uncoated copper grid (#400). We added elastase and 50 $\mu\text{M ATP}$ to the droplet on the grid, and washed the grid with assay buffer. After addition of the microtubule suspension, the grid was washed with 5 mM Na-phosphate (pH 7.5) and negatively stained with 1% uranyl acetate³⁰. Electron micrographs were taken with a JEOL 2000-EX microscope.

Received 2 March; accepted 31 March 1998.

1. Brokaw, C. J. Operation and regulation of the flagellar oscillator. *Cell Movement* **1**, 267–279 (1989).
2. Shingyoji, C., Murakami, A., & Takahashi, K. Local reactivation of Triton-extracted flagella by iontophoretic application of ATP. *Nature* **265**, 269–270 (1977).
3. Takahashi, K., Shingyoji, C., & Kamimura, S. Microtubule sliding in reactivated flagella. *Soc. Exp. Biol. Symp.* **35**, 159–177 (1982).
4. Shingyoji, C. & Takahashi, K. Cyclical bending movements induced locally by successive iontophoretic application of ATP to an elastase-treated flagellar axoneme. *J. Cell Sci.* **108**, 1359–1369 (1995).
5. Kamimura, S. & Kamiya, R. High-frequency nanometre-scale vibration in 'quiescent' flagellar axonemes. *Nature* **340**, 476–478 (1989).
6. Smith, E. F. & Sale, W. S. Regulation of dynein-driven microtubule sliding by the radial spokes in flagella. *Science* **257**, 1557–1559 (1992).
7. Svoboda, K., Schmidt, C. F., Schnapp, B. J. & Block, S. M. Direct observation of kinesin stepping by optical trapping interferometry. *Nature* **365**, 721–727 (1993).
8. Higuchi, H., Muto, E., Inoue, Y. & Yanagida, T. Kinetics of force generation by single kinesin molecules activated by laser photolysis of caged ATP. *Proc. Natl Acad. Sci. USA* **94**, 4359–4400 (1997).
9. Gibbons, I. R. Dynein ATPases as microtubule motors. *J. Biol. Chem.* **263**, 15837–15840 (1988).
10. Sale, W. S. & Satir, P. Direction of active sliding of microtubules in *Tetrahymena* cilia. *Proc. Natl Acad. Sci. USA* **74**, 2045–2059 (1977).
11. Fox, L. A. & Sale, W. S. Direction of force generated by the inner row of dynein arms on flagellar microtubules. *J. Cell Biol.* **105**, 1781–1787 (1987).
12. Vale, R. D. & Toyoshima, Y. Y. Rotation and translocation of microtubules *in vitro* induced by dyneins from *Tetrahymena* cilia. *Cell* **52**, 459–469 (1988).
13. Sale, W. S. & Fox, L. A. Isolated β -heavy chain subunit of dynein translocates microtubules *in vitro*. *J. Cell Biol.* **107**, 1793–1797 (1988).

14. Yokota, E. & Mabuchi, I. C/A dynein isolated from sea urchin sperm flagellar axonemes. Enzymatic properties and interaction with microtubules. *J. Cell Sci.* **107**, 353–361 (1994).
15. Johnson, K. A. Pathway of the microtubule-dynein ATPase and the structure of dynein: a comparison with actomyosin. *Annu. Rev. Biophys. Chem.* **14**, 161–188 (1985).
16. Svoboda, K. & Block, S. M. Force and velocity measured for single kinesin molecules. *Cell* **77**, 773–784 (1994).
17. Ishijima, A. *et al.* Single-molecule analysis of the actomyosin motor using nano-manipulation. *Biochem. Biophys. Res. Commun.* **199**, 1057–1063 (1994).
18. Finer, J. T., Simmons, R. M. & Spudis, J. A. Single myosin molecule mechanics: piconewton forces and nanometre steps. *Nature* **368**, 113–119 (1994).
19. Goodenough, U. & Heuser, J. Structural comparison of purified dynein proteins with *in situ* dynein arms. *J. Mol. Biol.* **180**, 1083–1118 (1984).
20. Sale, W. S., Goodenough, U. W. & Heuser, J. E. The substructure of isolated and *in situ* outer dynein arms of sea urchin sperm flagella. *J. Cell Biol.* **101**, 1400–1412 (1985).
21. Wang, Z., Khan, S. & Sheetz, M. Single cytoplasmic dynein molecule movements: characterization and comparison with kinesin. *Biophys. J.* **69**, 2011–2023 (1995).
22. Thomas, N. & Thornhill, R. A. The physics of biological molecular motors. *J. Phys. D* **31**, 253–266 (1998).
23. Thomas, N. & Thornhill, R. A. Possible origin of tension oscillations in muscle preparations. *J. Physiol. (Lond.)* **491**, 124P (1996).
24. Moss, A. G., Sale, W. S., Fox, L. A. & Witman, G. B. The α subunit of sea urchin sperm outer arm dynein mediates structural and rigor binding to microtubules. *J. Cell Biol.* **118**, 1189–1200 (1992).
25. Moss, A. G., Gatti, J.-L. & Witman, G. B. The motile β /IC1 subunit of sea urchin sperm outer arm dynein does not form a rigor bond. *J. Cell Biol.* **118**, 1177–1188 (1992).
26. Omoto, C. K. Mechanochemical coupling in eukaryotic flagella. *J. Theor. Biol.* **137**, 163–169 (1989).
27. Shimizu, T., Marchese-Ragona, S. P., & Johnson, K. A. Activation of the dynein adenosinetriphosphate by cross-linking to microtubules. *Biochemistry* **28**, 7016–7021 (1989).
28. Gee, M. A., Heuser, J. A. & Vallee, R. B. An extended microtubule-binding structure within the dynein motor domain. *Nature* **390**, 636–639 (1997).
29. Gibbons, B. H. & Gibbons, I. R. The effect of partial extraction of dynein arms on the movement of reactivated sea-urchin sperm. *J. Cell Sci.* **13**, 337–357 (1973).
30. Katayama, E. The effects of various nucleotides on the structure of actin-attached myosin subfragment-1 studied by quick-freeze deep-etch electron microscopy. *J. Biochem.* **106**, 751–770 (1989).

Acknowledgements. We thank K. Takahashi and R. A. Thornhill for discussion and critical reading of the manuscript, and E. Muto and Y. Inoue for suggestions about the preparation of microtubules. This work was supported by a Kurata Research Grant and a Grant from Nissan Science Foundation (to C.S.).

Correspondence and requests for materials should be addressed to C.S. (e-mail: chikako@biol.s.u-tokyo.ac.jp).

KNOW YOUR COPY RIGHTS RESPECT OURS

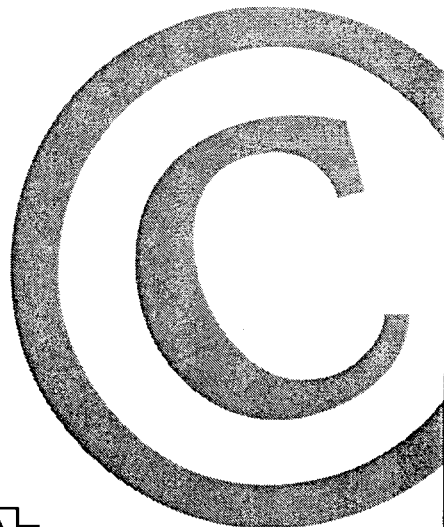
The publication you are reading is protected by copyright law. This means that the publisher could take you and your employer to court and claim heavy legal damages if you make unauthorised photocopies from these pages.

Photocopying copyright material without permission is no different from stealing a magazine from a newsagent, only it doesn't seem like theft.

The Copyright Licensing Agency (CLA) is an organisation which issues licences to bring photocopying within the law. It has designed licensing services to cover all kinds of special needs in business, education, and government.

If you take photocopies from books, magazines and periodicals at work your employer should be licensed with CLA.

Make sure you are protected by a photocopying licence.



The Copyright Licensing Agency Limited
90 Tottenham Court Road, London W1P 0LP
Telephone: 0171 436 5931
Fax: 0171 436 3986

A tale of two cultures

Cell and tissue culture are the focus of this week's lineup which includes synthetic culture media, an automated tissue resistance measurement system, a cell culture respirometer and trace elements for media enhancement.

Protocol

From Synoptics

A new colony counter designed to reduce plate counting workloads by up to 80 per cent

Using image analysis technology and a new colony counting algorithm, this system counts plates quickly and features control software that produces results in formats that are appropriate to microbiologists. Results are acquired, stored and reported in formats compliant with good laboratory practice. There are dilution series and plate production features, including spiral plating techniques. The system offers automatic detection of colonies, and is said to eliminate problems due to unevenly poured plates and variations in background density between different plates in batch. Performance is enhanced by recording the count data immediately and directly in a computer, eliminating the possibility of transcription errors. The colony counter can handle fluorescing bacteria or nuclei when used in conjunction with a microscope and is claimed to be equally effective with all types of media.

Reader Enquiry No. 100

REMS AutoSampler

From World Precision Instruments

An automated PC-controlled microplate tissue resistance measurement system

Designed for use with cell culture microplates, this system is built to provide speed, precision and contamination-free measurement of tissue resistance. It should prove useful in determining the effect of drugs on membrane transport, measuring resistance up to 20 k Ω in 24- and 96-well microplates. Microplates are placed on a fluid-filled base plate that also serves as the bottom half of a transmembrane electrode pair. The x-y-z movement of the robot precisely positions the probe electrode over a sample well. A Windows95-compatible software package controls tissue resistance measurements.

Reader Enquiry No. 101

RM 2145 microtome

From Leica

An ergonomically designed semi-motorized rotary microtome

All key microtome functions are located on an ergonomic control panel, separate from the main unit. The setup increases comfortable use through an adjustable inclination angle and easy-to-read LED display. Selected



A cut above: Leica's RM 2145 microtome.

section and trimming thicknesses are displayed separately. A quick-release clamping system can be operated with one hand for fast changeover of specimen clamps. To ensure safety, a 70-mm vertical stroke is employed, and there is an anti-tilt function within the specimen orientation system. In addition to the conventional, full handwheel rotation for sectioning, the unit features the new Ergomode rocking motion. In this mode, the handwheel is moved back and forth in small increments, and the instrument detects all changes in the direction of rotation, automatically translating them into feed or retraction. Other features include a knife holder with lateral displacement function and an integrated, section-waste tray that transfers waste directly from the knife into the tray.

Reader Enquiry No. 102

Cell culture respirometer

From Columbus Instruments

A system for the quantitative measurement of cell growth, utilizing O₂ consumption and CO₂ production of small cell culture samples

The respirometer can be used for measuring tissue and bacterial cultures, as well as oxidation of enzymes. Measurements can be conducted in ambient air or other gas atmospheres, as well as at different temperatures and over time frames from hours to weeks. Computing the amount of O₂ consumed and CO₂ produced is accomplished by measuring changes in gas concentration in the headspace above the cell culture. Gas sensors do not contact the cell/tissue culture and each cell chamber is isolated from the equipment by fine bacterial/cell filters. These culture chambers can be any size, and 1–80 samples can be measured simultaneously. Owing to the high sensitivity of the system (0.2 μ l O₂ h⁻¹), oxygen consumption of small samples can be measured with high accuracy. Optional gas sensors allow the measurement of other gases, such as H₂, H₂S or CH₄. The

unit is controlled by a PC, providing periodic numerical results and graphs.

Reader Enquiry No. 103

Cytoperm 2

From Heraeus

A contamination-resistant CO₂ incubator

This incubator features total sterilization capability, providing protection for cell cultures. The 180 °C, hot-air sterilization cycle is sufficient to kill the types of contamination normally encountered in cell culture applications. Double-level protection for cultures is ensured by combining a pyrolytic germ barrier with a screen of six individual doors to the incubator's interior. The pyrolytic germ barrier works by passing water (stored externally) through a fire furnace at 500 °C, so that puffs of sterile steam are sent into the incubator to achieve the desired humidity. The unit features HEPA-filtered CO₂ input, a 220-l work chamber and an air-jacketed heating system.

Reader Enquiry No. 104

Media and supplies

Trace elements

From Mediatech

Trace elements for serum-free cell culture

The use of trace elements may reduce or eliminate the need for growth factors in certain serum-free applications. Three trace element concentrates are offered to assist customers with these applications. The supplements can be used individually or in combinations, with the choice of trace element solutions depending on the type of cell work being done and the basal medium being used. For example, the Formulation A may be applicable to cells known to respond positively to the presence of selenium, provided the presence of copper, zinc or iron is tolerated or beneficial.

Reader Enquiry No. 105

Synthetic culture media

From Autogen Bioclear

Bio Media's range of synthetic culture media, which contains a stable form of L-glutamine

This stable dipeptide (L-Ala-L-Glu) replaces the normal L-Glu found in the company's original formulations. This stable form is said to increase the shelf life of the media to 24 months when stored at 4 °C. In addition, they are stated to give improved results in the culture of hybridomas for monoclonal antibody production. Five types of pre-formulated culture media are currently available with the

stabilized form of L-Glu: RPMI 1640, DMEM Dulbecco's, Ham F12, DMEM/F12 mixture 1:1 and Dulbecco's Iscove Modified.

Reader Enquiry No. 106

ES-Cult product line

From StemCell Technologies

For in vitro haematopoietic differentiation and maintenance of mouse (ES) cells

This is a two-step *in vitro* differentiation procedure for embryonic stem (ES) cells in methylcellulose-based media. This can be employed as an alternative to the generation of chimaeric and knockout mice for gene function studies, particularly when the knockouts are lethal to embryos. This product line includes sera, liquid and methylcellulose-based media, supplements and accessories. All are pre-tested for their utility in the various stages of ES cell differentiation experiments. ES cells are totipotent cells derived from the inner cell mass of day 3.5 blastocysts. They can be maintained *in vitro* for extended periods without loss of their capacity to contribute to all cell lineages when reimplanted into a blastocyst.

Reader Enquiry No. 107

Tissue culture products

From Greiner

A complete line of flasks, tubes, dishes and multi-well plates for tissue culture

As well as offering traditional cell culture products, Greiner also has a parallel range of products designed specifically for suspension culture use. The suspension culture grade also provides optimum cell growth with maximum yield, but without cell adherence, reducing mechanical or enzymatic damage. Cellstar flasks have 25-cm², 75-cm² and 175-cm² growth areas and are offered with standard and vented caps to provide contamination-free gas exchange. The screw caps incorporate a 0.2-µm hydrophobic membrane filter so that cultures grown in these flasks can be safely vented without risk of bacterial or fungal contamination. Also included in this range are roller bottles for culturing higher volumes of cells in the same physical dimensions of a standard roller bottle.

Reader Enquiry No. 108

Bacti Caps

From Clark Scientific

A simple method of ensuring positive culture tube closure with controlled gas interchange

This range is suitable for all standard test tube sizes and is designed to fit rimless glass and plastic tubes with an outside diameter of 13, 16, 19, 25 and 38 mm. They are made from a chemically resistant polypropylene formulation, protecting them from attack by culture media, and most acids, alkalis, alcohols and

esters. Three flexible fins are moulded on the inner wall of each cap to ensure a positive fit, even when minor variations in tube sizes occur. Lugs are moulded into their base to prevent a hermetic seal with the top of the tube. The position of these lugs permits a controlled gas interchange. The caps can be repeatedly autoclaved at 121 °C (108 kPa steam pressure), although they are not suit-

able for dry heat sterilization. There are 12 colours for coding purposes and these caps are supplied in packs of 100, in assortments of sizes and colours.

Reader Enquiry No. 109

These notes are compiled by Brendan Horton from information provided by the manufacturers. For more details, fill in the reader service card bound inside the journal.

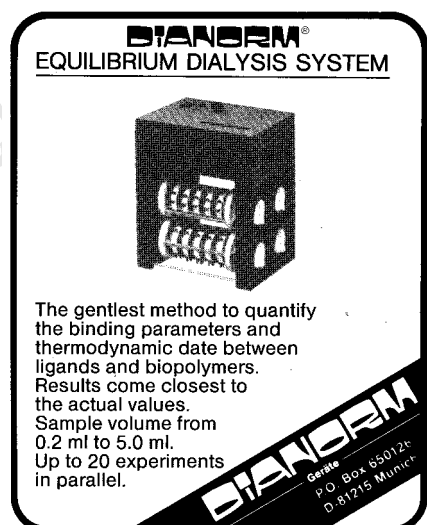


serotec
...better products, better service, better quality...

Antibodies to Human CD's	Cytokine & Growth Factor Reagents
Antibodies to Mouse CD's	Cytoskeletal Compound Reagents
Antibodies to Rat CD's	Flow Cytometry Reagents
Antibodies to Ovine CD's	Holone Film Reagents
Antibodies to Bovine CD's	Cell Signalling Reagents
Veterinary Research Antibodies	Neurological Reagents
Antibodies to Hormones	Reagents for Immunology
Antibodies to Microbial Antigens	Reagents for Biotechnology
Apoptosis Reagents	TCS Reagents
Adhesion Molecule Products	Tissue Culture Products

FOR RESEARCH USE ONLY
Serotec, 22 Watlington, Watlington, Oxford OX9 1B, UK
Tel: +44 (0)1865 852700 Fax: +44 (0)1865 852701
Technical Enquiries Tel: +44 (0)1865 852713
Order Numbers Tel: +44 (0)1865 852723
www.serotec.co.uk

READER ENQUIRY NO. 23

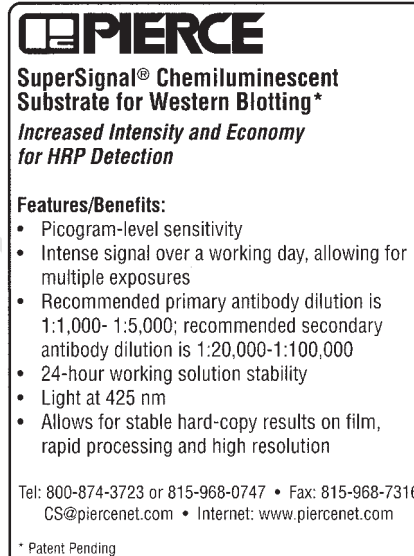


BIANORM®
EQUILIBRIUM DIALYSIS SYSTEM

The gentlest method to quantify the binding parameters and thermodynamic data between ligands and biopolymers. Results come closest to the actual values. Sample volume from 0.2 ml to 5.0 ml. Up to 20 experiments in parallel.

BIANORM
Geräte
P.O. Box 65012
D-81215 Munich

READER ENQUIRY NO. 61



PIERCE
SuperSignal® Chemiluminescent Substrate for Western Blotting*
Increased Intensity and Economy for HRP Detection

Features/Benefits:

- Picogram-level sensitivity
- Intense signal over a working day, allowing for multiple exposures
- Recommended primary antibody dilution is 1:1,000-1:5,000; recommended secondary antibody dilution is 1:20,000-1:100,000
- 24-hour working solution stability
- Light at 425 nm
- Allows for stable hard-copy results on film, rapid processing and high resolution

Tel: 800-874-3723 or 815-968-0747 • Fax: 815-968-7316
CS@piercenet.com • Internet: www.piercenet.com
* Patent Pending

READER ENQUIRY NO. 33

ANTIBODIES
For Research & Routine Histology

Inflammation – Angiogenesis – Tumors
Leukocytes – Endothelial Cells
Pathological Markers

Custom Monoclonals and Polyclonals

BMA Biomedicals Ltd.
(800) 634-3183 (US)
Int'l: +41 61 811 62 22
+41 61 811 60 06 (fax)
info@bma.ch

READER ENQUIRY NO. 28

PACEMAKER®

Quality assured
BS:EN ISO9002
accredited

Custom peptides, conjugates and antibodies
Unrivalled experience and expertise with assured quality

For more information on AFFINITY's Peptide and Protein Research Services, please contact:

AFFINITY Research Products Ltd., Mamhead Castle, Exeter, EX6 8HD, UK
Tel: +44/0 1626 891010; Fax: +44/0 1626 891090; URL: <http://www.affinity-res.com>

READER ENQUIRY NO. 32